50C

NAVAL RESEARCH LOGISTICS QUARTERLY

OFFICE OF NAVAL RESEARCH

CONTENTS

ARTICLES

"Hedging" on Statistical Assumptions 207
Lionel Weiss

Budget Constraints in Logistics Models 215
G. Hadley and T. M. Whitin

Optimal Disposal Policies 221
Yoichiro Fukuda

Decision on Retention of Excess Stock Following a Normal Probability Law of Obsolescence and Deterioration 229
C. Mohan and R. C. Garg

On the Timing of Development Expenditures and The Retirement of Military Equipment 235
A. S. Manne

On the Estimation of Scale Parameters 245
Lionel Weiss

A Transportation Algorithm and Code 257
Merrill M. Flood

Approximately Optimal One-Dimensional Search Policies in Which Search Costs Vary Through Time 277
Brian Gluss

A Delivery-Lag Inventory Model With an Emergency Provision (The Single-Period Case) 285
E. W. Barankin

RECENT PUBLICATIONS 313

VOL. 8, NO. 3 SEPTEMBER 1961

NAVEXOS P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Rear Admiral H. E. Eccles, USN (Retired)
The George Washington University

O. Morgenstern
Princeton University

F. D. Rigby
Office of Naval Research

D. M. Gilford
Office of Naval Research

Commander J. V. Rosapepe, SC, USN
Managing Editor
Office of Naval Research
Washington 25, D. C.

B. J. Beall
Secretary
Office of Naval Research
Washington 25, D. C.

ASSOCIATE EDITORS

R. Bellman, RAND Corporation
J. C. Busby, Jr., Commander, SC, USN
J. S. Campbell, Technical Operations Inc.
W. W. Cooper, Carnegie Institute of Technology
J. G. Dean, Captain, SC, USN
G. Dyer, Vice Admiral, USN (Retired)
P. L. Folsom, Captain, USN
M. A. Geisler, RAND Corporation
A. J. Hoffman, General Electric Company
S. Karlin, Stanford University
H. W. Kuhn, Princeton University
J. Laderman, Office of Naval Research, Br. Office
W. H. Marlow, The George Washington University
R. E. McShane, Vice Admiral, USN (Retired)
C. A. Messenheimer, Captain, SC, USN
W. F. Millson, Captain, SC, USN (Retired)
M. I. Rosenberg, Captain, USN (Retired)
D. Rosenblatt, The George Washington University
H. A. Sachaklian, Colonel, USAF
E. K. Scofield, Captain, SC, USN
J. R. Simpson, Bureau of Supplies and Accounts
J. S. Skoczylas, Colonel, USMC
H. Solomon, The George Washington University
I. Stakgold, Northwestern University
E. D. Stanley, Captain, SC, USN
C. Stein, Jr., Captain, SC, USN (Retired)
R. M. Thrall, University of Michigan
C. B. Tompkins, University of California
J. D. Wilkes, Office of Naval Research

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. Subscription Price: $1.50 a year in the U.S. and Canada, $2.00 elsewhere; $0.50 for a single copy. Reprints of an individual article can be purchased in quantities of 100 or more. Requests for the purchase price of reprints of a particular article should be sent to the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C.

The views and opinions expressed in this quarterly are those of the authors and not necessarily those of the Office of Naval Research.

The issuance of this publication approved by the Secretary of the Navy on 24 March 1961.

Permission has been granted to use the copyrighted material appearing in this publication.

"HEDGING" ON STATISTICAL ASSUMPTIONS*†

Lionel Weiss

Cornell University

It is pointed out that in many cases statistical assumptions, such as the assumption of a normal population, really cannot be known to be true with the absolute certainty that the word "assumption" has come to imply. In fact, in some instances data collected for purposes other than testing the assumptions cast serious doubt on the validity of the assumptions. A simple safety or hedging device is proposed, and illustrated for the case where the problem is to estimate the q^{th} quantile (m, say) of a distribution assumed to be normal with given standard deviation σ. Let $\overline{X}$, R denote the sample mean and the q^{th} sample quantile, respectively. Under the assumption, the best estimate of m is $\overline{X} + t\sigma$, where t is a properly chosen constant, while R is a reasonable estimate of m even if the assumption is not true. The hedging device is to use $\overline{X} + t\sigma$ as the estimate if $|\overline{X} + t\sigma - R| \leq d$, and to use R as the estimate if $|\overline{X} + t\sigma - R| > d$, where d is a preassigned value. In order to choose d to reflect our faith in the validity of the assumption, and to obtain the operating characteristics of our estimation rule, it is necessary to know the joint distribution of $\overline{X}$ and R. This is shown to approach a bivariate normal distribution as the sample size increases, for any population satisfying mild regularity conditions.

INTRODUCTION

In conventional statistical terminology, an "assumption," as distinguished from a "hypothesis," is something that is assumed to be true and is not intended to be subjected to testing. However, in many statistical problems, assumptions are made which obviously cannot be known to be true with absolute certainty. An example would be the common assumption of normality of various physical populations. This assumption is seldom the result of analysis of observations, and an attempt to use the central limit theorem to "prove" normality is often dubious because of such factors as lack of independence.

Sometimes when data are collected for purposes other than testing our assumptions, they cast serious doubt on the assumptions made. Then there is a question of what to do: continue as though the assumptions were true, or change the assumptions to conform to the data? If we adopt the latter course, we may have to collect new data to check our new assumptions, for the dangers of allowing data to suggest assumptions and then using the same data are well known. It is the purpose of this paper to point out that under some dircumstances we can legitimately change our assumptions on the basis of data and still use the same data for our ultimate purpose. Broadly speaking, we examine the data to see if they seriously contradict our assumptions, and if they do we switch from the decision that would have been appropriate

*Manuscript received November 27, 1959.
†Research under contract with the Office of Naval Research.

617404 O - 61 - 1

under the assumptions to the decision that is appropriate under weaker assumptions that our data do not contradict. Such a procedure typically requires the examination of the joint distribution of a pair of random variables: one random variable representing the decision appropriate under the assumptions, the other the decision appropriate under weaker assumptions.

ESTIMATING A POPULATION QUANTILE

Suppose that our problem is to estimate the q^{th} quantile of a continuous population, where q is a given value in the open interval (0, 1). The loss incurred is the squared difference between our estimate and the true value of the population quantile. We are going to take n independent identically distributed observations from the population. We denote the arithmetic mean of the n observations by T, and the q^{th} quantile of the sample by R. The q^{th} sample quantile is the sample value with exactly [nq] sample values smaller than it, where [nq] is the largest integer not greater than nq. Here we are ignoring the possibility of ties among the sample values.

If we assume that the population is normal with unknown mean but with known standard deviation σ, then the minimax estimate of the population quantile is $T + t\sigma$, where t satisfies the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left\{-\frac{1}{2}y^2\right\} dy = q .$$

Under much weaker assumptions, the appropriate estimate of the population quantile is R. If the assumption of normality with standard deviation σ fails to hold (because of the wrong value for the standard deviation, or nonnormality of the population), the use of $T + t\sigma$ as our estimate can be disastrous if the true population is very different from the assumed population. However, we can protect ourselves against this possibility by the following "hedging" device: if $|T + t\sigma - R| \leq d$, then we estimate the population quantile by $T + t\sigma$, while if $|T + t\sigma - R| > d$, we estimate the population quantile by R. Here d is a positive predetermined constant. The logic of this device is clear: if $|T + t\sigma - R| > d$, it is evident that our assumption of normality with standard deviation σ is false, and therefore we use an estimate that is reasonably good when our assumption is false.

A natural question is: "Why not simply use R as an estimate, if we are not sure that our assumption is true?" The answer is that if we have a lot of faith in our assumption, but faith stopping short of absolute certainty, the hedging device gives us a chance to use the optimal estimate under the assumption, abandoning it only if the data (which have to be collected anyway) cast serious doubt on the validity of the assumption. The stronger our faith in the assumption, the greater will be the value we choose for d. Thus the choice of d involves the same subjective elements as the choice of a level of significance when testing a hypothesis.

In order to choose d in a way that accurately reflects our faith in the assumption, as well as to analyze the operating characteristics of the estimation procedure incorporating the hedging device, it is clearly necessary to study the joint probability distribution of T and R. We do this in the next section, for large sample size n.

THE LIMITING JOINT DISTRIBUTION OF T AND R

It is well known that under mild restrictions on the population distribution, when T and R are properly standardized their marginal distributions approach normal distributions as n increases (see [1], page 215 and page 369). It would be surprising if the limiting joint distribution were not bivariate normal, but nobody seems to have proved the limiting normality of the bivariate distribution. We prove the limiting normality in the following theorem.

THEOREM: Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables, with common cumulative distribution function $F(x)$. We assume that $F(x)$ is absolutely continuous, and denote its derivative by $f(x)$, and we also assume that the distribution has a finite variance σ^2. We denote the population mean by μ. We assume that the population has a unique q^{th} quantile, which we denote by m, and that $f(m) > 0$ and $f(x)$ has a continuous derivative in a neighborhood of m. T, R denote the sample arithmetic mean and the q^{th} sample quantile, respectively, where the sample consists of the values $X_1, \ldots, X_n$. Then as n increases, the joint distribution of

$$\frac{\sqrt{n}\,(T - \mu)}{\sigma},$$

$$\frac{\sqrt{n}\,f(m)\,(R - m)}{\sqrt{q(1 - q)}}$$

approaches the bivariate normal distribution with zero means, unit variances, and correlation coefficient equal to

$$\frac{\sqrt{q(1 - q)}}{\sigma}\left(\frac{\int_m^\infty x\,f(x)\,dx}{1 - q} - \frac{\int_{-\infty}^m x\,f(x)\,dx}{q}\right).$$

Proof: We denote by S_1 the sum of the $[nq]$ sample values which are smaller than R, and by S_2 the sum of the $n - [nq] - 1$ sample values which are greater than R. Then $T = (S_1 + R + S_2)/n$. The joint conditional distribution of S_1, S_2 given R is that of two independent random variables, S_1 having the marginal distribution of a sum of $[nq]$ independent random variables, each with density function $f(x)/F(R)$ if $x < R$ and zero if $x \geq R$, S_2 having the marginal distribution of a sum of $n - [nq] - 1$ independent random variables, each with density function $f(x)/(1 - F(R))$ if $x > R$ and zero if $x \leq R$. We denote the expressions

$$\frac{S_1 - [nq]\,\dfrac{\int_{-\infty}^{R} x\,f(x)\,dx}{F(R)}}{\sqrt{[nq]\left[\dfrac{\int_{-\infty}^{R} x^2 f(x)\,dx}{F(R)} - \left(\dfrac{\int_{-\infty}^{R} x\,f(x)\,dx}{F(R)}\right)^2\right]^{\frac{1}{2}}},$$

$$\frac{S_2 - (n - [nq] - 1)\,\dfrac{\int_{R}^{\infty} x\,f(x)\,dx}{1 - F(R)}}{\sqrt{n - [nq] - 1}\left[\dfrac{\int_{R}^{\infty} x^2 f(x)\,dx}{1 - F(R)} - \left(\dfrac{\int_{R}^{\infty} x\,f(x)\,dx}{1 - F(R)}\right)^2\right]^{\frac{1}{2}}}$$

by Y_1, Y_2 respectively. As n increases, the joint conditional distribution of Y_1, Y_2 given R approaches the joint distribution of two independent standard normal variables.

Next we denote $\sqrt{n}\,f(m)\,(R - m)/\sqrt{q(1-q)}$ by Z, so that $R = m + Z\sqrt{q(1-q)}/\sqrt{n}\,f(m)$. We then have

$$F(R) = q + Z\sqrt{q(1-q)}/\sqrt{n} + o(Z/\sqrt{n}),$$

$$\int_{-\infty}^{R} x\,f(x)\,dx = \int_{-\infty}^{m} x\,f(x)\,dx + Zm\sqrt{q(1-q)}/\sqrt{n} + o(Z/\sqrt{n}),$$

$$\int_{-\infty}^{R} x^2 f(x)\,dx = \int_{-\infty}^{m} x^2 f(x)\,dx + Zm^2\sqrt{q(1-q)}/\sqrt{n} + o(Z/\sqrt{n}),$$

$$\int_{R}^{\infty} x\,f(x)\,dx = \int_{m}^{\infty} x\,f(x)\,dx - Zm\sqrt{q(1-q)}/\sqrt{n} + o(Z/\sqrt{n}),$$

$$\int_{R}^{\infty} x^2 f(x)\,dx = \int_{m}^{\infty} x^2 f(x)\,dx - Zm^2\sqrt{q(1-q)}/\sqrt{n} + o(Z/\sqrt{n}).$$

Expressing S_1, S_2 in terms of Y_1, Y_2 respectively, and using the expressions developed in the preceding paragraph, we find after some manipulation that $(S_1 + S_2 - n\mu)/\sqrt{n}$ is equal to

$$Z\sqrt{q(1-q)}\left(\frac{\int_m^\infty x\,f(x)\,dx}{1-q} - \frac{\int_{-\infty}^m x\,f(x)\,dx}{q}\right)$$

$$+ Y_1\left[\int_{-\infty}^m x^2 f(x)\,dx - \frac{\left(\int_{-\infty}^m x\,f(x)\,dx\right)^2}{q} + o(Z/\sqrt{n})\right]^{\frac{1}{2}}$$

$$+ Y_2\left[\int_m^\infty x^2 f(x)\,dx - \frac{\left(\int_m^\infty x\,f(x)\,dx\right)^2}{1-q} + o(Z/\sqrt{n})\right]^{\frac{1}{2}} + o(Z/\sqrt{n})\,.$$

It follows immediately that the limiting conditional distribution of $(S_1 + S_2 - n\mu)/\sqrt{n}$ given R is normal, with mean equal to

$$Z\sqrt{q(1-q)}\left(\frac{\int_m^\infty x\,f(x)\,dx}{1-q} - \frac{\int_{-\infty}^m x\,f(x)\,dx}{q}\right)$$

and variance equal to

$$\int_{-\infty}^\infty x^2 f(x)\,dx - \left[\frac{\left(\int_{-\infty}^m x\,f(x)\,dx\right)^2}{q} + \frac{\left(\int_m^\infty x\,f(x)\,dx\right)^2}{1-q}\right].$$

Since we have that $\sqrt{n}\,(T-\mu)/\sigma$ is equal to $(S_1 + S_2 - n\mu)/\sigma\sqrt{n} + R/\sigma\sqrt{n}$, we have that the limiting conditional distribution of $\sqrt{n}\,(T-\mu)/\sigma$ given R is normal, with mean

$$\frac{Z\sqrt{q(1-q)}}{\sigma}\left(\frac{\int_m^\infty x\,f(x)\,dx}{1-q} - \frac{\int_{-\infty}^m x\,f(x)\,dx}{q}\right),$$

and variance

$$\frac{\int_{-\infty}^\infty x^2 f(x)\,dx}{\sigma^2} - \frac{\left(\int_{-\infty}^m x\,f(x)\,dx\right)^2}{q\sigma^2} - \frac{\left(\int_m^\infty x\,f(x)\,dx\right)^2}{(1-q)\sigma^2}.$$

Since the marginal distributions of Z and of $\sqrt{n}\,(T-\mu)/\sigma$ are known to approach standard normal distributions as n increases, the theorem follows immediately.

APPLICATIONS

The theorem can be applied to find the value of d that will give a desired probability of using $T + t\sigma$ as the estimate of the population quantile, when the assumption of normality with standard deviation σ is true and we are using the hedging device described in the second section. Suppose we want the probability of using $T + t\sigma$ as the estimate when the population distribution is actually normal with standard deviation σ to be equal to α, α a predetermined value. Then we must have $P[|T + t\sigma - R| \leq d] = \alpha$, the probability being evaluated under the assumption that the population distribution is normal with standard deviation σ. In this case the q^{th} quantile m is equal to $\mu + t\sigma$, and we have

$$P[|T + t\sigma - R| \leq d] = P\left[-\sqrt{n}\, d \leq \sigma\left\{\frac{\sqrt{n}\,(T - \mu)}{\sigma}\right\} - \frac{\sqrt{q(1-q)}}{f(m)}\left\{\frac{\sqrt{n}\, f(m)\,(R - m)}{q(1-q)}\right\} \leq \sqrt{n}\, d\right].$$

But for large n, the random variable in the center of this last inequality has approximately a normal distribution with mean zero and variance V (say) equal to

$$\sigma^2 + \frac{q(1-q)}{f^2(m)} - \frac{2q(1-q)}{f(m)}\left(\frac{\int_m^\infty x\, f(x)\, dx}{1-q} - \frac{\int_{-\infty}^m x\, f(x)\, dx}{q}\right).$$

by the theorem of the preceding section. Here f(x) is the normal density function with mean μ and standard deviation σ, and an easy computation gives $V = \sigma^2\,[2\pi q(1-q)\exp\{t^2\} - 1]$. Therefore, it follows that as n increases,

$$\frac{\sqrt{n}\, d}{\sigma\sqrt{2\pi q(1-q)\exp(t^2) - 1}}$$

must approach the value c given by

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{c} \exp\left(-\frac{1}{2}y^2\right) dy = \frac{1}{2}(1 + \alpha).$$

As a numerical example, suppose that $q = 3/4$, so it is the third quantile that is to be estimated, and $\alpha = 0.95$. Then $t = 0.6745$, $c = 1.96$, so that for large n, d is approximately equal to $1.813\,\sigma/\sqrt{n}$.

Another application of the theorem of the section just mentioned gives the large-sample operating characteristic of our hedging estimation procedure. Let Y denote the random variable defined as follows: $Y = T + t\sigma$ if $|T + t\sigma - R| \leq d$, $Y = R$ if $|T + t\sigma - R| > d$. Our loss when using the hedging estimation procedure is then equal to $(Y - m)^2$. In any case where the population distribution function satisfies the hypotheses of the theorem, a straightforward though somewhat lengthy calculation gives

$$\lim_{n\to\infty} P\left[n(Y - m)^2 \leq b\right] \equiv \lim_{n\to\infty} P\left[-b^{\frac{1}{2}} \leq \sqrt{n}\,(Y - m) \leq b^{\frac{1}{2}}\right],$$

for any positive b. Tables of the bivariate normal distribution must be used.

OTHER PROBLEMS

The problem discussed was one of the simplest imaginable. More complicated ones can easily be imagined. For example, while assuming that our observations are independent and identically distributed, we may want to hedge against the possibility that our observations contain a time trend.

There are other methods of hedging. One way would be to perform an all-purpose test of fit such as the chi-square test to test the hypothesis that the population is normal with standard deviation σ. If the hypothesis is accepted, use $T + t\sigma$ as the estimate, otherwise use R. We may be able to show that one method of hedging is better than another, in some reasonable sense.

REFERENCE

[1] H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, Princeton (1951).

* * *

BUDGET CONSTRAINTS IN LOGISTICS MODELS*

G. Hadley†

University of Chicago

T. M. Whitin†

University of California, Berkeley

In application of operations research models in the area of logistics, important constraints are typically encountered in the form of limitations on the funds appropriated. A supply system must be operated within whatever budgeted funds are appropriated by Congress. In the present article, various possible types of budget constraints are discussed as well as their impact on certain types of operations research models. The possible misallocations that can result from the setting of specific budgets designated for particular purposes are discussed. The confusion existing in the literature concerning the type of constraints that are actually encountered versus those that are imposed upon the mathematical models is pointed out as well as the typical inconsistency of budget constraints and steady-state models. The difference between a steady-state budget and a transition-phase budget is clarified.

INTRODUCTION

Attention is focussed on procurement budgets and transportation budgets as typical of the budget problems encountered. In practice, the general policy is that funds allocated for transportation cannot be used for procurement and vice versa. It should be fairly obvious at the outset that additional flexibility can be gained by having these funds interchangeable, since larger procurement expenditures can affect reductions in transportation expenditures and conversely large expenditures on transportation can reduce the level of inventories and hence reduce the total quantity procured.

The budgets under consideration are specified in terms of actual dollars to be spent. In other words, opportunity costs or costs such as stockout costs or inventory carrying charges are not the type that are actually budgeted since they are not real out-of-pocket costs. Procurement costs consist of fixed costs independent of the number of units purchased plus variable costs which are a function of the number of units ordered. Part of the fixed cost of an order consists of the costs of placing procurement orders. However, typically these order processing costs are not charged to the procurement budget which includes only those funds actually spent at the source on procurement.

*Manuscript received September 12, 1960.

†The authors are indebted to the research and development branch of Bureau of Supplies and Accounts, U. S. Navy, and Captain C. H. Stein of United Research, Inc., for sponsoring this work.

In a similar manner, many of the costs associated with transportation costs are not charged to the transportation budget. For example, transportation costs from the source are sometimes charged to the procurement budget. The fixed costs of analyzing and processing redistribution requests are not usually charged in the transportation, although such costs frequently exceed the funds paid directly to carriers.

Further ambiguities concerning the appropriate budget to charge various expenditures against arise because of the time lag between the placing of orders and the receipt of materials. Here possibilities may exist for charging the cost of the order to the procurement budget in either the fiscal year in which the order is placed or the fiscal year when the materials are delivered. The year in which the order is placed is the normal procedure and will be assumed to be the procedure used for purposes of this article.

The existence of budget constraints can, under certain circumstances, lead to substantial modification in the results of operations research models. The following paragraphs set forth some of the problems and illustrate the way in which budget constraints may be handled under varying assumptions concerning the nature of these constraints.

AVERAGE PROCUREMENT BUDGET

Many of the inventory models discussed in the literature are based upon steady-state assumptions.[1] For example, the models involving the determination of economical procurement quantities and reorder point quantities and those involving the determination of optimal review cycles are steady-state models. If the steady-state assumptions are not to be violated, the average amount procured per year must be equal to the average amount demanded. This is true of each item of the system as well as for the system as a whole. If less than the average amount demanded is procured, then, on the average, the stocks will steadily decrease. When more than the average demand is procured, the stocks will, on the average, steadily increase.

Now suppose that a budget constraint is imposed on the average amount to be spent per year on procurement. It will be imagined that the budget is flexible so that in any given year any amount required by the random demand may be spent in such a manner that the average amount spent will be a constraint. Such a budget may be specified for a given item or for a whole group of items. It is more realistic to have the budget specified for a whole group of items. However, before considering this case, it is instructive to examine the case where the budget constraint is applied to just a single item.

In the event that the total cost of the item is independent of the number of items procured, then it is meaningless to impose a budget constraint because if the average budget is not CY where C is the total unit cost and Y the average demand, then the steady state assumptions are violated. For items other than low-demand items the assumption of constant unit cost is fairly realistic, since the costs of processing order is not charged to the procurement budget. There is therefore a large group of items for which there is only one value of the average budget consistent with the steady-state assumptions.

For some items, however, the total unit cost of the item changes with the number of units purchased because of fixed set up costs and/or economies or diseconomies of scale. In this case the cost per unit C(Q) depends on Q, the number of units ordered, and the average

[1] For example, see K. J. Arrow, T. Harris, and J. Marschak, "Optimal Inventory Policy," Econometrica (July, 1951), pp. 250-72; T. M. Whitin and J. W. T. Youngs, "A Method for Calculating Optimal Inventory Levels and Delivery Time," Naval Research Logistics Quarterly September 1955, pp. 157-73.

cost of procurement per year is YC(Q). In this event it is sometimes possible to vary the budget while still maintaining an equilibrium model. The smallest average budget which will yield an equilibrium model is that for which on the average Y units per year in quantities Q* where C(Q*) is the lowest unit price possible. It is important to note that when the optimal Q is altered by budget considerations the expected total costs including stockout costs and the costs of carrying inventory are higher, even though average procurement expenditures are somewhat reduced.

Consider next the case where an average procurement budget, designated by B, is specified for a whole group of items. First, the average budget requirements for those items whose unit price is independent of quantity can be subtracted from B to give $\hat{B}$. For the remaining items, then minimum average budgets are summed. If the sum is greater than $\hat{B}$, it is not possible to maintain equilibrium. If this sum is less than $\hat{B}$, it is possible to maintain a steady-state model. If $\hat{B}$ exceeds the sum of the average procurement costs for quantities determined by an optimizing model with no budgetary constraint imposed, then the average budget is greater than that necessary to operate the system optimally. If the sum of the procurement costs for the optimal quantities exceeds $\hat{B}$, then the problem is to allocate $\hat{B}$ among the items in an optimal way. The logical criterion for doing this is to allocate $\hat{B}$ so as to minimize the sum over all the items of the expected total costs, including stockout costs and the costs of carrying inventory. If the subscript j refers to item j, it is desired to minimize $\Sigma\, TC_j$ subject to the constraint $\Sigma\, Y_j C_j (Q_j) \leq B$. This problem may be solved in theory by the use of Lagrange multipliers.

ABSOLUTE PROCUREMENT BUDGET

In the previous section the case was examined where an average procurement budget was to be set which might be different from the optimal one. It was observed that the system could be operated for a range of budgets less than the optimal one. However, in such cases the total system costs including stockout and carrying costs was higher than for the optimal system. Note that only an average budget was referred to. Random fluctuations in demand will cause more than this to be spent in some years and less to be spent in other years. The budget of any given year was varied as required to meet actual procurement costs, only the average value of the constraint being specified. In the present section, it is assumed that an absolute budget is specified each year and no more than this amount can be spent. This budget may be the same each year or it may be varied from year to year. The problem now takes on a fundamentally different nature.

If the same amount is budgeted each year and if the amount is in the range of average budgets discussed in the previous section, which allow the steady-state operation of the system, one might attempt to use the budget in the manner discussed above for the average budget. However, such attempts must be unsuccessful. In those years when procurements turn out to be less than the budgeted amount the money would not be spent and would be lost forever since budget money cannot be carried over from one year to the next. Such a situation presents at least two difficulties. The losses from budget money in certain years might be serious enough to cause the system to run down hill and destroy the steady-state assumption. Another realistic problem is that if the budget money is not completely spent in a given year there is a very good chance that it will be cut for the following year.

This points out that when an absolute budget constraint is imposed on a year to year basis, a steady-state model is no longer suitable in general. This is even more clear where

the budget is specified and may vary from year to year. An entirely different and exceedingly complicated model is required.

An absolute procurement budget can become a serious problem during the transition from the existing system to a new type of steady-state system operation, particularly if larger safety stocks are needed in the new system. Additional procurement funds must be allocated to effect the transition. These additional funds must be sufficient to procure the additional safety stocks. It is important to note that the additional funds are needed only for the transition phase and that procurements in subsequent years involve only the purchase on the average of the average quantity demanded per year. On the other hand, if the safety stocks in the new system are lower than for the existing system, procurement expenses may be reduced during the transition phase. Again the change is only in the transition phase. The manner in which the transition phase is carried out is rather arbitrary. For example, if additional procurement funds are needed, they may be made available all at once or over a period of years.

In addition to the increase in safety stocks required for transition to the new steady state, problems may arise if the average value of working stocks, say $\sum_j Q_j^1 C_j/2$ for the new steady state exceeds that of the old.[2] For the first year, the average increase in procurement expenditure would be $\sum_j \left(Q_j^1 - Q_j\right)\frac{C_j}{2}$. Although the additional stock will be used in subsequent years, nevertheless the budget problem could be critical in the first year. Fortunately, however, it is possible to phase the transition in such a manner that the problem may be reduced. The amount of procurement expenditures in the first year is $\Sigma Y_j C_j$ + value of final inventory (on hand plus on order) — value of initial inventory (on hand plus on order). If the transition is phased in such a manner that expected year-end inventories are at a low level, the procurement expenditures for the first year are reduced.

TRANSPORTATION BUDGET CONSTRAINTS

In addition to procurement budget constraints, it is frequently true in the operation of a supply system that transportation budget constraints are also imposed. As noted previously, only funds paid to carriers are charged to the transportation budget. It was also observed previously that it is desirable from the point of view of flexibility to have a combined procurement and transportation budget rather than separate budgets. Of course there may be disadvantages to such a combined budget from other points of view. For example, there may be administrative difficulties associated with attempting to control the funds if more than a single group makes decisions about using them. However, if decisions concerning procurement and transportation are made centrally at a control point, this should not prove to be a problem. An overly stringent budget limitation on transportation expenses can lead to unnecessarily expensive interim procurement to avoid transportation costs or to stock-out costs higher than are necessary. A combined budget would help to eliminate imbalances between the transportation and procurement budgets.

Transportation budgets (and procurement budgets also) which do not charge the fixed costs of processing an order to the budget foster the shipping of unnecessarily small quantities. A more reasonable but more complicated system from the accounting point of view is to charge all costs associated with transportation to the transportation budget. As was the case for

[2] Q_j^1 represents the purchase quantity of item j in the new steady state.

procurement budgets, either an average or an absolute transportation budget may be set. In the latter case, a tight transportation budget at year-end can lead to a large number of shipments of much less than economical size.

The level of the transportation budget has no effect on the steady-state nature of supply models, unlike procurement budgets. As the transportation budget is reduced, less redistribution can be tolerated in the system. If an increase in stockouts is not to result, safety stock levels must be increased or recourse taken to additional local procurements. Either of these alternatives requires an increase in the procurement budget. Thus, there exist trade-offs between procurement and transportation budgets. It is interesting to note that although this trade-off plays an important role in the operation of a logistics system, it has been essentially ignored in most operations research models in the logistics field. Not a single published model includes both the fixed and variable transportation cost components or determines analytically the proper balance between procurement and transportation budgets.

OTHER APPROACHES TO THE CONSTRAINT PROBLEM

Several other approaches to budget constraints have been suggested in previous work. Generally these approaches appear to be unsatisfactory in one way or another. In one report,[3] an attempt was made to impose a constraint on the expected sum of procurement expenditures over all future time. This approach is not reasonable in practice, since Congress does not set budgets according to the present worth of all future expenditures, but rather on an annual basis. Another criticism of the method proposed is that it involves estimations of obsolescence that would be extremely difficult to make.

A more recent[4] approach consisted of minimizing an expression for expected total costs (procurement costs plus inventory carrying charges plus stockout costs), and of varying the stockout cost so as to bring the average holding costs plus average fixed procurement costs to some given level. This was accomplished through the use of a common multiplicative parameter applied to the stockout costs. The procedure described is equivalent to minimizing the fixed ordering costs plus inventory carrying costs subject to a specified average system essentiality weighted unit weeks of shortage constraint. The equivalence follows immediately from the theory of Lagrange multipliers. There is nothing wrong with this procedure if its objectives are indeed those that one seeks to achieve. However, there seems to be little relationship between the procedure and realistic budget constraints. The technique of varying the cost of a unit week of shortage for an item of unit essentiality in such a way as to bring the average sum of fixed procurement costs plus carrying charges to a given level is not too meaningful for several reasons. First, carrying charges are not part of a procurement budget. Secondly, the fixed procurement charges considered are those incurred by the Navy in processing an order and these are not ordinarily charged to a procurement budget either. Thus, no costs which are actually charged to a procurement budget are included in this procedure. Furthermore, the technique deals only with average quantities. In any given year the value of fixed ordering costs plus carrying charges may vary over a wide range. Finally, it is not clear that the value assigned to the cost of a unit week of shortage for an item of unit essentiality, i.e., the Lagrange multiplier value, bears any relationship to the actual costs of being out of stock.

---

[3] Stock Control Policy, Aviation Supply Office, Stanford Research Institute, 1958.
[4] Advanced Stock Control Procedures, Electronics Supply Office, Stanford Research Institute, January 1959.

Two other procedures which are not realistic from the procurement budget point of view, but which are as interesting as the methods discussed above are as follows. First one might attempt to minimize the expected value of fixed procurement costs plus stockout costs subject to a constraint requiring that the average inventory be a given value. The theory of Lagrange multipliers shows that this is equivalent to minimizing the sum of fixed ordering costs, plus holding costs, plus stockout costs with the holding cost being the Lagrange multiplier. Thus, the holding cost could be varied until the average inventory is brought to the desired level. The other procedure is to minimize the expected value of the sum of holding costs plus stockout costs subject to the constraint that the average number of procurements per year is a given value. This is equivalent to minimizing the sum of fixed procurement, holding, and stockout costs, with the Lagrange multiplier being the cost of placing an order. By varying the cost of placing an order, the average number of orders per year placed can be brought to the desired value.

CONCLUSIONS

The task of developing models which include budget constraints in a realistic way seems to be difficult. This article has pointed out some of the problems. The important difference between average and absolute budget has been emphasized. It has also been pointed out that steady in state models which minimize the expected value costs it is only possible to impose average budget constraints and that the type of constraints which can be imposed usually have no relation to any physical quantities actually budgeted. The realistic problem of how to deal with absolute budgets appears to be extremely difficult and it seems that no attempts have been made to solve this problem. It has also been observed that because of the peculiarities in the way charges are made to procurement and transportation budgets in actual practice, that decisions can be fostered at various operating levels which are detrimental to the operation of the entire military supply system.

* * *

OPTIMAL DISPOSAL POLICIES*†

Yoichiro Fukuda

Planning Research Corporation
Los Angeles, California

In this paper we analyze the inventory model in which the disposal of surplus items is taken into consideration as well as the purchase of stock in anticipation of future demand. Our main concern is to formulate the optimal ordering and disposal policy in a simple form, and it is assumed that the decision to be made at the beginning of each period is always one of the following: (1) ordering, (2) disposal, (3) no ordering and no disposal. The multiple installation model is treated by the method of implied penalty described by Clark and Scarf. The simple optimal policies are explicitly determined under various sets of assumptions on the cost structure of the inventory model.

INTRODUCTION

In this paper we analyze the multiechelon inventory model in which the disposal of surplus items is taken into consideration as well as the purchase of stock in anticipation of future demand. Our main concern is to determine the optimal ordering and disposal policy in the simple form, and it is assumed that the decision to be made at the beginning of each period is always one of the following: (1) ordering, (2) disposal, (3) no ordering and no disposal.

The system to be considered in the following is only of the type

$$[\quad]_N \rightarrow [\quad]_{N-1} \rightarrow [\quad]_{N-2} \cdots [\quad]_2 \rightarrow [\quad]_1 .$$

and it is treated by the method of implied penalty described by Clark and Scarf [1], and incorporates suggestions made by Scarf through informal correspondence with Planning Research Corporation.

There are four kinds of costs operative in this inventory model: an ordering cost depending on the number of items to be purchased; a disposal cost depending on the amount to be disposed (this may be a negative cost indicating that revenue may be obtained from disposal); a holding cost depending on the stock on hand; a shortage cost, proportional to the excess of demand over available stock during the period. An important assumption here is that any excess demand is to be backlogged. The holding cost and shortage cost are taken as linear functions; therefore, if the stock on hand at the beginning of the period is x, then the expected cost during the period, exclusive of ordering and disposal costs, is given by

*Manuscript received June 17, 1960.
†This paper is based on the report PRC R-152, Planning Research Corporation, prepared for Bureau of Supplies and Accounts, Department of the Navy, under Contract Nonr-2713(00).

$$L(x) = \begin{cases} p \int_0^\infty (\xi - x)\, \phi(\xi)\, d\xi , & x \le 0 , \\ hx + p \int_x^\infty (\xi - x)\, \phi(\xi)\, d\xi , & x > 0 , \end{cases}$$

where h and p are the unit holding and shortage costs, respectively, and the quantity demanded during the period is represented by a random variable ξ with the density function $\phi(\xi)$.

The ordering cost is assumed to be directly proportional to the amount to be purchased, and it is also assumed that the stock is delivered one time period after it is ordered. The disposal cost is given by a unit disposal cost times the amount to be disposed, and the unit disposal cost may be negative, indicating that revenue may be obtained from disposed items. Another assumption to be made concerning the disposal of surplus items is that the items to be disposed are withdrawn from the installation at the beginning of each period. There may be different ways of handling the items to be disposed (for example, to retain them in the installation until the end of the period), but we restrict ourselves to this assumption in this paper.

If there is made a decision to dispose of some items at some lower echelon, then they are immediately withdrawn from the echelon; but one period time length is necessary for them to reach the next higher echelon. In the similar manner, it takes one period time length to get stock down from the higher echelon. The only way of getting stock out of the system completely is to move it step by step up the echelons until it reaches the highest echelon and gets out of it.

In addition to the implied shortage penalty charged by the lower echelon, as in the case of no disposal, an intermediate echelon in the present problem must suffer from a penalty charged by the upper echelon owing to its disposal policy. In view of this complexity, the second echelon of a three echelon system is analyzed in detail in the next section.

DETERMINATION OF THE OPTIMAL POLICIES

As usual, we start with an analysis of the one period model. It is clear that no ordering is made in this model; hence, we go directly to the problem of determining the optimal disposal level. Designating by $f_j^k(x)$ the minimum total expected costs for the jth echelon over the k periods in the future with x as the amount of stock at the beginning of the present period, we set up the following equation for $x > 0$:

$$f_2^1(x) = \min_{0 \le v \le x} \left\{ d_2(x - v) + L_2^1(v) + \Lambda_2^1(x) \right\} . \tag{1}$$

Here, d_2 represents the unit disposal cost for the echelon two, and $L_2^1(v)$ represents the one-period costs of operation for this echelon which is given as follows:

$$L_2^1(x) = p_2 \int_0^\infty (\xi - x)\, \phi_2^1(\xi)\, d\xi \,, \qquad x \le 0 \,,$$

$$= h_2 x + p_2 \int_x^\infty (\xi - x)\, \phi_2^1(\xi)\, d\xi \,, \qquad x > 0 \,.$$

$\Lambda_2^1(x)$ is the penalty charged by the echelon three to the echelon two on the basis that if the echelon two holds more items than the optimal disposal level of the echelon three, then the echelon three is obligated to dispose its stocks only to the same level as the echelon two, hence, this incurs some extra costs to the echelon three, which should be charged to the echelon two as penalty. If we denote the optimal disposal level of the echelon three for this period by v_3^1, then $\Lambda_2^1(x)$ is given by

$$\Lambda_2^1(x) = 0 \,, \qquad x \le v_3^1 \,,$$

$$= -d_3(x - v_3^1) + L_3^1(x) - L_3^1(v_3^1) \,, \qquad x > v_3^1 \,.$$

An important thing to be noticed here is that Λ_2^1 is independent of the optimal disposal level of v_2^1 of the echelon two, since whatever amount to be disposed out of the echelon two (or the installation two) does not reach the installation three until the next period, and therefore does not affect the decision at the installation three (or the echelon three) in the present period. We have no shortage penalty charged by the echelon one in this period because of the fact that no ordering is made in any echelon.

To determine the optimal disposal level v_2^1, we obtain from (1) the following expression

$$H_2^1(v) = -d_2 v + L_2^1(v) \,.$$

Since $L_2^1(v)$ is convex in v, there exists a unique $v_2^1 > 0$ such as $H'^1_2(v_2^1) = -d_2 + L'^1_2(v_2^1) = 0$, if we have

(2) $$-d_2 + h_2 - p_2 < 0 \,, \qquad \text{and} \qquad -d_2 + h_2 > 0 \,.$$

If we have $x > v_2^1$, then we dispose $x - v_2^1$, and if $x \le v_2^1$, we dispose nothing. Therefore, $f_2^1(x)$ is given as follows:

$$f_2^1(x) = L_2^1(x) + \Lambda_2^1(x) \,, \qquad x \le v_2^1 \,,$$

$$= d_2(x - v_2^1) + L_2^1(v) + \Lambda_2^1(x) \,, \qquad x > v_2^1 \,,$$

$$f'^1_2(x) = L'^1_2(x) + \Lambda'^1_2(x) \,, \qquad x < v_2^1 \,,$$

$$= d_2 + \Lambda'^1_2(x) \,, \qquad x > v_2^1 \,.$$

617404 O - 61 - 2

We notice that $f_2^1(x)$ is convex in x and $f'^1_2(x) \to d_2 - d_3 + h_3 > 0$ as $x \to \infty$.

Now we proceed to the two period model. We argue that the ordering would be made at least for a very small stock at the beginning of the period under a reasonable set of assumptions, and we set, for a small x

$$f_2^2(x) = \min_{w \geq x} \left\{ c_2(w - x) + L_2^2(x) + K_2^2(x) + \Lambda_2^2(x) \right.$$

$$\left. + \alpha \int_0^\infty f_2^1(w - \xi)\, \phi_2^2(\xi)\, d\xi \right\}.$$

A very significant assumption to be made in the above formulation is that the penalty $\Lambda_2^2(x)$ should be independent of the optimal ordering level w_2^2. If the installation three should try to fulfill the ordering from the installation two first, and then decide how much to dispose from the remaining stock, then this assumption does not hold. However, if it should dispose first, and then try to take care of the ordering, as much as it could, then the assumption is valid, since the implied shortage penalty is charged to the echelon three in case any unfilled ordering exists in the echelon two.

K_2^2 represents the penalty to be charged by the echelon one in case of unfilled ordering, and this is apparently independent of the ordering in the present period. Because of these considerations, the quantity necessary to determine the optimal ordering level w_2^2 turns out to be

$$G_2^2(w) = c_2 w + \alpha \int_0^\infty f_2^1(w - \xi)\, \phi_2^2(\xi)\, d\xi \ .$$

Since $f_2^1(w)$ is convex in x, there exists a unique $w_2^2 > 0$ such as

$$G'^2_2(w_2^2) = c_2 + \alpha \int_0^\infty f'^1_2(w_2^2 - \xi)\, \phi_2^2(\xi)\, d\xi = 0 ,$$

if we have

$$c_2 - \alpha p_2 < 0 . \tag{3}$$

The optimal ordering policy is to order up to w_2^2 if $x < w_2^2$, and not to order if $x \geq w_2^2$. Next, we set for $x \geq w_2^2$,

$$f_2^2(x) = \min_{x \geq v \geq 0} \left\{ d_2(x - v) + L_2^2(v) + K_2^2(v) + \Lambda_2^2(x) \right.$$

$$\left. + \alpha \int_0^\infty f_2^1(v - \xi)\, \phi_2^2(\xi)\, d\xi \right\}.$$

As stated in the one stage model, $\Lambda_2^2(x)$ is independent of the disposal level, but $K_2^2(v)$ is not, since the stock level is immediately reduced to the desired level before any ordering from the echelon one is taken care of. Hence, the quantity to be used to determine the optimal disposal level v_2^2 is represented by

$$H_2^2(v) = - d_2 v + L_2^2(v) + K_2^2(v) + \alpha \int_0^\infty f_2^1(v - \xi)\, \phi_2^2(\xi)\, d\xi \ .$$

and $K_2^2(v)$ is given by

$$K_2^2(x) = c_1(x - w_1^2) + \alpha \int_0^\infty \left\{ f_1^1(x - \xi) - f_1^1(w_1^2 - \xi) \right\} \phi_1^2(\xi)\, d\xi \qquad x < w_1^2 \ ,$$

$$= 0 \ , \qquad x \geq w_1^2 \ .$$

We notice that $K_2^2(v)$ is dependent on w_1^2, and, hence, we must have the values of w_1^2, w_2^2, w_3^2 before we start calculating the values of v_1^2, v_2^2, v_3^2. This is entirely feasible, since we only need $f'^1_1(x)$, $f'^1_2(x)$, and $f'^1_3(x)$ to determine w_1^2, w_2^2, w_3^2. If we recapitulate here the sequence of calculations to be made in determining the optimal levels, we first determine v_1^1, v_2^1, and v_3^1. Using these values, $f_1^1(x)$, $f_2^1(x)$, and $f_3^1(x)$ are specified, and hence, $f'^1_1(x)$, $f'^1_2(x)$, and $f'^1_3(x)$ are available. They are sufficient to determine w_1^2, w_2^2, and w_3^2, which are, in turn, instrumental in specifying $K_2^2(x)$, $K_3^2(x)$. Using $K_2^2(x)$ and $K_3^2(x)$, v_1^2, v_2^2, and v_2^3 are determined, and $f_1^2(x)$, $f_2^2(x)$, and $f_3^2(x)$ are completely specified. From here on, the process repeats the similar cycle of calculations.

Returning to the determination of v_2^2, we remark that $H_2^2(v)$ is convex in v, therefore, there exists a unique $v_2^2 > 0$ such as

$$H'^2_2(v_2^2) = - d_2 + L'^2_2(v_2^2) + K'^2_2(v_2^2) + \alpha \int_0^\infty f'^1_2(v_2^2 - \xi)\, \phi_2^2(\xi)\, d\xi = 0 \ ,$$

if we have $H'^2_2(0+) < 0$ and $H'^2_2(\infty) > 0$. This is indeed so, since we have

$$H'^2_2(0+) = d_2 + h_2 - p_2 + c_1 - \alpha p_1 - \alpha p_2 = - d_2 + h_2 - (1 + \alpha)p_2 + c_1 - \alpha p_1 < 0$$

by (2) and (3), and

$$H'^2_2(\infty) > - d_2 + h_2 > 0$$

by (2).

Another important thing here is that we should have $v_2^2 \geq w_2^2$. If we check $H'^2_2(w_2^2)$, we have

$$H'^2_2(w_2^2) = - d_2 + L'^2_2(w_2^2) + K'^2_2(w_2^2) + \alpha \int_0^\infty f'^1_2(w_2^2 - \xi)\, \phi_2^2(\xi)\, d\xi$$

$$= - d_2 - c_2 + L'^2_2(w_2^2) + K'^2_2(w_2^2) \leq - d_2 - c_2 + L'^2_2(w_2^2)$$

$$< - d_2 - c_2 + h_2 .$$

Hence, if we have another assumption

$$- d_2 - c_2 + h_2 \leq 0 \tag{4}$$

then we have $H'^2_2(w_2^2) < 0$ and this is sufficient to assure $w_2^2 < v_2^2$. To obtain the explicit form of $f_2^2(x)$, we need the expression of $\Lambda_2^2(x)$ which is given by

$$\Lambda_2^2(x) = 0 , \qquad x \leq v_3^2 ,$$

$$= - d_3(x - v_3^2) + L_3^2(x) - L_3^2(v_3^2) + K_3^2(x) - K_3^2(v_3^2)$$

$$+ \alpha \int_0^\infty \left\{ f_3^1(x - \xi) - f_3^1(v_3^2 - \xi) \right\} \phi_2^2(\xi)\, d\xi , \qquad x > v_3^2 .$$

Now $f_2^2(x)$ is specified as follows:

$$f_2^2(x) = c_2(w_2^2 - x) + L_2^2(x) + K_2^2(x) + \Lambda_2^2(x)$$

$$+ \alpha \int_0^\infty f_2^1(w_2^2 - \xi)\, \phi_2^2(\xi)\, d\xi , \qquad x < w_2^2 ,$$

$$= L_2^2(x) + K_2^2(x) + \Lambda_2^2(x) + \alpha \int_0^\infty f_2^1(x - \xi)\, \phi_2^2(\xi)\, d\xi , \qquad w_2^2 \leq x \leq v_2^2 ,$$

$$= d_2(x - v_2^2) + L_2^2(v_2^2) + K_2^2(v_2^2) + \Lambda_2^2(x)$$

$$+ \alpha \int_0^\infty f_2^1(v_2^2 - \xi)\, \phi_2^2(\xi)\, d\xi , \qquad x > v_2^2 ,$$

$$f'^{2}_{2}(x) = -c_2 + L'^{2}_{2}(x) + K'^{2}_{2}(x) + \Lambda'^{2}_{2}(x), \qquad x < w^{2}_{2},$$

$$= L'^{2}_{2}(x) + K'^{2}_{2}(x) + \Lambda'^{2}_{2}(x) + \alpha \int_{0}^{\infty} f'^{1}_{2}(x - \xi)\, \phi^{2}_{2}(\xi)\, d\xi, \qquad w^{2}_{2} < x < v^{2}_{2},$$

$$= d_2 + \Lambda'^{2}_{2}(x), \qquad x > v^{2}_{2}.$$

We notice that $f^{2}_{2}(x)$ is convex in x and $f'^{2}_{2}(x) \to d_2 - d_3 + h_3(1 + \alpha) > 0$ as $x \to \infty$. Thus, the necessary induction steps are completed and we have the similar results for the general n stage model.

REFERENCE

[1] Clark, A. J., and Scarf, H., "Optimal Policies for a Multi-Echelon Inventory Problem," Management Science, Vol. 6, No. 4 (July 1960), pp. 475-490.

* * *

DECISION ON RETENTION OF EXCESS STOCK FOLLOWING A NORMAL PROBABILITY LAW OF OBSOLESCENCE AND DETERIORATION*

C. Mohan and R. C. Garg

Naval Research Group, Defence Science Laboratory, Delhi

This paper gives a method by which economic retention quantity of a material having a known shelf-life and following a normal probability law of obsolescence and deterioration has been calculated. The relevant equations are solved by trial and error method.

INTRODUCTION

Every stocking agency is usually faced with the problem of assessing the quantity of material which can be retained on economic grounds, if at any particular instant the stock appears to be in excess of its demands. Simpson [1] has derived a formula for taking disposal decisions assuming a constant probability of obsolescence and deterioration. In this article a normal probability law of obsolescence and deterioration has been considered for all materials having a known shelf-life. The cumulative probability of obsolescence for the two cases plotted against time is shown in Figure 1. Economic retention period has then been arrived at by comparing the two costs, one due to expenditure incurred on storage plus the losses due to obsolescence and deterioration and the other due to the cost of purchasing the same stock again some years hence. An example has been worked out giving the economic retention quantity for a material of shelf-life of 10 yr and standard deviation 1.5 for decision-making points at intervals of 2 yr. The same method has been applied in the case of very short shelf-life items, such as dry batteries and films. In this case, a large stock in excess of the demand has to be held to meet unexpected demands which might arise due to unforeseen commitments. A disposal decision becomes vital here on economic grounds before the expiry of the shelf-life. A table has been worked out for this category of stores which gives the economic retention periods for dry batteries of shelf-life 9 months and standard deviation 1.5 months when the disposal decisions are made at an interval of one month.

THEORY

In deriving the formula the following factors have been assumed:

1. The annual demand of the agency throughout remains known and is constant.
2. The future price for procurement of the same item remains constant.
3. Other costs of the system like the cost of storage and the rate of interest also remain constant.

*Manuscript received September 26, 1960.

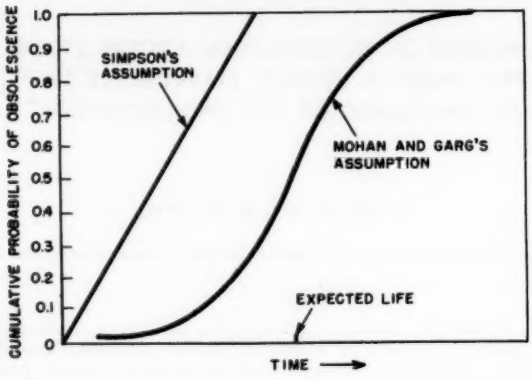

Figure 1 - The cumulative probability of obsolescence for Simpson's assumption and Mohan and Garg's assumption, plotted against time

4. The cost of the material is its unit price V at the time of taking decision; that its disposal value if sold is represented as a fraction D of its unit price V. There is a finite probability that its value will drop to zero in any one year because the material becomes completely obsolete or deteriorates in storage.

5. The obsolescence and deterioration follows a normal probability law with mean 'm' and a known standard deviation 's,' where 'm' is the expected shelf-life of the material. It may be observed that the probability of not becoming obsolete in the first x years is $(1 - F_x)$, where

$$F_x = \frac{1}{s\sqrt{2x}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x - m}{s}\right)^2} dx .$$

Again the conditional probability of obsolescence and deterioration for the interval, $(x, x + dx)$ is

$$Z_{(x)}\, dx = \frac{f(x)}{1 - F_x}\, dx$$

where $f(x)$, $Z_{(x)}$ are the probability densities of obsolescence and deterioration and conditional obsolescence and deterioration, respectively. The conditional probability density is an increasing function of x which reflects the expectation that obsolescence and deterioration increase with time because of ageing.

DEFINITION OF VARIABLES

U = expected average annual demand

X = N/U - the number of years for which a stock N will meet expected demand

V = value of the material, assumed to be standard price per unit at the time of taking decision

D = fraction of unit price of material at the time of taking decision which will be realized in disposal sales

i = interest rate

r = annual storage cost of material, expressed as a fraction of the rupee value of material, e.g., $r = 0.01$ means that it costs 0.01 Re to store material worth one rupee for 1 yr

C_R = Total cost of retention of stock

C_D = Total disposal cost

ECONOMIC RETENTION QUANTITY

Consider the costs which will be incurred under two alternatives of retaining or disposing the excess material.

The storage costs C_R consist of two types of costs, (i) the storage cost, (ii) the loss in value if the material becomes worthless.

The storage costs are taken as r rupees per year. However, there exist probabilities $F_1, F_2 - F_1, \ldots, F_x - F_{x-1}$ that the material will become worthless by obsolescence and deterioration and have to be thrown away in any one of the x successive years. In such a situation, storage cost will no longer be incurred. Therefore, the storage cost should be reduced each year. This probability may be expressed as $(1 - F_x)$. If the material had not been stored, the money used for storage costs would have been invested at compound interest and so the cost of storage must include the interest otherwise earned. Therefore, for the period of x years under consideration, the storage costs are

$$r\left[(1 - F_1)(1 + i)^x + (1 - F_2)(1 + i)^{x-1} + \ldots\ldots\ldots + (1 - F_x)(1 + i)\right].$$

Also the loss in value if the material should become worthless is $1 - (1 - F_x)$. Therefore,

$$(1)\quad C_R = 1 - (1 - F_x) + r\left[(1 - F_1)(1 + i)^x + (1 - F_2)(1 + i)^{x-1} + \ldots\ldots\ldots + (1 - F_x)(1 + i)\right].$$

Next we consider the cost of disposal of stocks. We may dispose of the material now and procure it again after x years. Therefore, the cost must be reduced by the return from the disposal sale D, and this amount should increase at compound interest rates for x years. Therefore,

$$(2)\quad C_D = 1 - D(1 + i)^x.$$

It is evident from (1) that C_R increases with x, on the other hand (2) shows that C_D decreases with x. At the break even point we should have

$$C_R - C_D = 0 . \tag{3}$$

At values of x less than that given by (3), the costs of disposing the stocks will be higher, at values greater than that given by (3), the costs of retaining the stock will be higher. x thus represents the economic retention period and $x \times U$ is the economic retention quantity, E.R.Q.

We now proceed to find the value of x from (3) with the help of (1) and (2).

SOLUTION OF THE EQUATION

Substituting the values of C_R and C_D from (1) and (2) in (3) we get

$$Db^x + r\left[(1 - F_1)\, b^x + (1 - F_2)\, b^{x-1} + \ldots\ldots\ldots + (1 - F_x)\, b\right] - (1 - F_x) = 0 , \tag{4}$$

where $b = 1 + i$.

Equation (4) gives the economic retention period if the decision is sought at the very start. If the decision is to be made at the nth year, Eq. (4) may be modified to:

$$Db^x + \frac{r}{(1 - F_n)}\left[(1 - F_{n+1})\, b^x + (1 - F_{n+2})\, b^{x-1} + \ldots\ldots\ldots + (1 - F_{n+x})\, b\right] - \left\{\frac{1 - F_{n+x}}{1 - F_n}\right\} = 0. \tag{5}$$

By giving different values to n in Eq. (5) and solving it by trial and error method, we obtain economic retention periods for different decision making years.

A numerical example is solved for which m = 10, s = 1.5, U = 100, r = 0.01 for various disposal values at decision taking years 0, 2, 4, 6, 8 and 10 and the results are given in Table 1.

TABLE 1

Results of the Numerical Example

Decision Year	D	E.R.P.	E.R.Q.
0	0.35	9.7	970
2	0.25	8.4	840
4	0.15	7.0	700
6	0.05	6.0	600
8	0.05	4.0	400
10	0.05	2.6	260

In the first few years the E.R.P. is very nearly equal to shelf-life. As the disposal value of the item goes down, the disposal decision assumes importance depending upon the size of stocks held in hand.

E.R.P. FOR SHORT SHELF-LIFE ITEMS

In the case of short shelf-life items like dry cells, batteries, and films the disposal decision is of interest in the last few months of the shelf-life of the item. A simple rate of interest i per month on the investment has been assumed for these cases. With this assumption the Eq. (5) becomes

$$D(1+ix)+\frac{r}{(1-F_n)}\left[(1-F_{n+1})(1+ix)+(1-F_{n+2})(1+i\overline{x-1})+\ldots\ldots\right.$$

(6)

$$\left.+(1-F_{n+x})(1+i)\right]-\left\{\frac{1-F_{n+x}}{1-F_n}\right\}=0\,,$$

where r is the monthly storage cost of material expressed as a fraction of the rupee.

Employing the method adopted in the previous section, Eq. (6) is solved for the case of batteries of shelf-life of 9 months with a standard deviation of 1.5 months, where r = 0.001, for various disposal values at different decision taking months, and the results are given in the Table 2.

TABLE 2

Results of the Adopted Method

Decision Month	D	E.R.P. in Months
1	0.85	6.00
2	0.85	5.00
3	0.85	4.00
4	0.50	4.75
5	0.10	5.80
6	0.10	4.80
7	0.10	3.80
8	0.10	2.80
9	0.10	2.30

The disposal decision regarding the optimum economic size of the stock which should be retained before its disposal value drops to zero is of importance in the last few months (from 5th to 9th month in this case) of its shelf-life. Thus, a short shelf-life item in store needs critical review of the stock position in the latter half of its shelf-life. As normal probability law holds good for most of these items, a disposal decision on economic grounds need be taken only once within the shelf-life of the item instead of reviewing the items every month.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to Dr. S. S. Srivastava, Director, Scientific Research (Navy), for his encouragement and valuable suggestions. We also wish to express our thanks to Major H. S. Subba Rao and Mr. Naunihal Singh for useful discussions during the preparation of this article.

REFERENCE

[1] Simpson, J. R., "A Formula for Decision on Retention or Disposal of Excess Stock," Naval Research Logistics Quarterly, 2:145-55 (1955).

* * *

ON THE TIMING OF DEVELOPMENT EXPENDITURES AND THE RETIREMENT OF MILITARY EQUIPMENT*

A. S. Manne†

Technical Military Planning Operation
General Electric Company‡
Santa Barbara, California

INTRODUCTION

This article is intended to yield insights in an exceedingly conjectural area—the timing of development expenditures and the retirement of military equipment. A highly simplified model of the equipment life-cycle is employed, a model in which the timing decisions are supposedly made in such a way as to maximize military effectiveness subject to a fixed budget limitation on the average annual rate of dollar expenditures. The tradeoffs considered here are threefold:

1. Spending dollars on development in order to obtain new types of equipment.
2. Spending dollars on procurement in order to obtain useable quantities of the new types.
3. Spending dollars for the operation of old and comparatively ineffective equipment.

If materiel is to be retired early and new types introduced frequently, then a large fraction of the budget must continually be spent on development, procurement, and training. Conversely, with late retirement and slow development, a large fraction must be spent on the costs of operating superannuated units.

Essentially, this model represents an extension to the military sphere of the "MAPI" technique devised originally by Terborgh to deal with the problems of industrial equipment replacement.[1] Unlike the MAPI formulation, this one is explicitly concerned with development costs—the costs involved in changing over from one state-of-the-art to its successor. The presence of development costs makes it necessary to move in a staircase pattern—to jump from one technology level to the next—instead of continually embracing each new opportunity for improvement.

Like the MAPI model, this stems from the ancient doctrine that no new piece of military equipment represents an "ultimate weapon." Instead, each new device constitutes just one more step in man's unending quest to perfect the means of destroying his fellow-men. From

*Manuscript received June 20, 1960.
†The author gratefully acknowledges his indebtedness to Harold Asher and Ernest Mosbaek, and hereby absolves them from any responsibility for the remaining defects.
‡This article was obtained from G.E.'s SP series. The SP series is designed to accommodate primarily speeches, papers and other presentations of individual staff members. Therefore, this document does not necessarily reflect a coordinated opinion of the Technical Military Planning Operation.
[1]G. Terborgh, Dynamic Equipment Policy, McGraw-Hill, 1949.

this viewpoint, there is no such thing as a stationary plateau of preparedness. Today's most fashionable weapon is always doomed to be out of style within a few short years.

Two specific conclusions emerge from this study: (1) that the optimal retirement age of military equipment is independent of the absolute scale of budgetary expenditures; and (2) that the optimal interval between successive generations within a weapons family depends upon the ratio of development costs to the size of the total budget, and not upon the absolute values of these parameters. Substantial economies of scale in equipment development are available to the country with the largest military budget.

In order to obtain specific results, this optimization model is based upon some admittedly oversimple assumptions. Of these assumptions, the most crucial is that at any point in time—within a given family of successive weapon systems—it is possible to construct quantitative relationships between the military effectiveness of the old technology and the new. In order to make use of the model in determining the optimum retirement age of successive missile systems, for example, it would be necessary to make some kind of crude estimate of the number of "old" technology Atlas squadrons required in order to be equivalent to a single "new" technology Minuteman squadron.

Once this substitution coefficient has been estimated through a comparison of systems effectiveness, the planner is then in a position to make use of the model proposed here for evaluating the optimal replacement age and/or the optimal interval between the development of new equipment types. It is the contention of this article that it is more appropriate for a military analyst to bring his judgment to bear upon the question of comparative effectiveness of military equipment than for him to be forced to come up with a purely intuitive solution to a still more difficult question: How should a limited budget be split between the competing demands for development, procurement, and operation?

ASSUMPTIONS

Assumption 1

The military planner is confronted with a fixed budgetary constraint on the average rate of annual expenditures. Subject to this constraint, he is free to choose both a retirement age for old equipment and the time-spacing between new models in such a way as to maximize military effectiveness. In this sense, the planner is free to vary the fractions of the over-all budget to be spent on development, on procurement, and on the operation of equipment.

Assumption 2

In order to get each successive new model into production, it will cost a lump-sum of $\underline{d}$ million dollars for that model. With $\underline{d}$ equal to \$50 million, for example, it will cost \$50 million yearly to develop a new model once each year. Only \$25 million yearly will be required; however, if new models are introduced once every two years. In a more sophisticated analysis, one might wish to express $\underline{d}$ as a function of the time elapsed since the development of the previous model. We will assume it to be a lump-sum independent of the elapsed time.

Assumption 3

For each dollar's worth of initial costs—excluding development, but including both procurement and training costs—involved in setting up a new military unit, it will take a constant annual total of $\underline{c}$ dollars in order to operate the equipment over its entire life.

For example, if $c = .2$, and if it costs \$100 million in order to set up a certain type of squadron, then the operating costs of this squadron will come to \$20 million during each of the years in which the unit is maintained in active condition.[2]

In view of the postulated proportionality between operating costs and initial costs, it will be convenient to normalize all of these calculations in terms of initial costs. Instead of counting military units in terms of wings or squadrons, it will be convenient to count in terms of, say, "megabucks," the amount of airpower which can be purchased in each successive generation at an initial cost of one million dollars. Each military unit defined in this way will entail subsequent operating costs of $\underline{c}$ million dollars during each year of its service life, $\underline{y}$.

Together, assumptions (2) and (3) mean that we have approximated a total cost "learning curve" with a linear function. That is, in order to procure a new type at the rate of $\underline{x}$ million-dollar military units annually over each of $\underline{z}$ years, the cumulated total of development plus initial costs spent upon that type will have amounted to $d + xz$ million dollars. The combined rate of these expenditures would then average out to $(d/z) + x$ millions yearly.

Assumption 4

The progress of military technology—both our own and that of an enemy—occurs at a comparatively steady rate over time. The best equipment that becomes available each year is continuously better than the best available during the preceding year. In other words, we ignore the possibility that technological breakthroughs do occur in reality, and that they are then followed by plateaus. Instead, both for analytical convenience and for simplicity in forecasting, it is assumed that technological progress occurs at a constant exponential rate. In mathematical terms, this consists of describing the military effectiveness of the best available technology of $\underline{t}$ years ago as equal to the fraction a^t in relation to a brand-new unit (see Figure 1). If, for example, a military systems analyst came to the conclusion that at each point in time, one-year-old military units of a certain type were only 80 percent as effective as the best available brand-new unit, this would amount to saying that the effectiveness factor, $\underline{a}$, was equal to .8. Furthermore, this would be taken to mean that a 2-yr-old unit was only 64 percent as effective as a brand-new unit, and so on.

Note that this formulation implies that the military effectiveness of a piece of equipment never quite drops to absolute zero. No matter what the magnitude of $\underline{t}$, a^t is always positive. $(0 \leq t \leq \infty;\ 0 < a < 1)$. It is implied that even an extra flintlock would have some military usefulness during a thermonuclear era. However, to adopt the policy of never retiring military equipment can only be optimal in a case where operating costs are negligible.

Assumption 5

Each expenditure of $\underline{d}$ million dollars upon development enables us to buy the best technology known at time $\underline{t}$. Figure 1, for example, is drawn up under the assumption of a steady procurement rate over time, two-year spacing between successive models, retirement at the end of six years, and again $a^t = .8^t$. This then would mean that at the beginning of each production run, the current effective worth of the newest one-third of the units in the inventory would each equal $(.8)^2 = .64$; of the middle-aged third, $(.8)^4 = .41$; and of the oldest third, $(.8)^6 = .26$. Under our policy of continuous procurement and retirement, the military

[2] If this model were constructed so as to take cognizance of either physical attrition of manned aircraft or else test firing of missiles, the annual operation costs would not then remain constant at a level of c, but would continuously decline during the service life of the equipment.

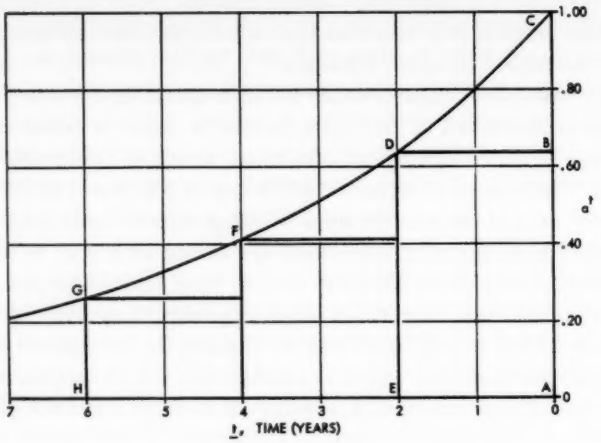

Figure 1 - a^t, comparative military effectiveness of equipment developed t year in the past

effectiveness of the force will be slightly higher during the middle of each run than at either the beginning or the end. Our measure of over-all effectiveness will be based upon the low point reached during each cycle.

Figure 1 suggests a direct geometrical representation of the distinction between expenditures on research and those on development. Through research, we hope to stimulate the production of useful new technology, and thereby steepen the slope of the comparative effectiveness curve. Through development, however, we are enabled to buy into the effectiveness curve at a given state-of-the-art in order to procure useable quantities of equipment. The decision to develop new models, say, once every 2 yr means that we consciously fail to reap the full benefits of technological advance. If it were economically feasible to keep abreast of the advances in technology by continuously developing new models of equipment, there would be no need to incur the losses of effectiveness associated with the triangular-shaped areas, BCD. The presence of development costs, however, makes it desirable to move in a staircase fashion—to jump from one technology plateau to the next—instead of moving along the curve GFDC in a continuous fashion. The degradation in effectiveness due to two-year spacing instead of continuous development of new models is given by the ratio of the two areas, ABDE/ABCDE.[3] Through closer spacing—and thereby higher development costs—it is possible to make the ratio of these areas approach unity.

This, then, constitutes the optimization problem: striking a reasonable balance between expenditures for development, procurement, and operation. Only through development expenditures can we fully capitalize upon advances in the basic technology. But only through procurement and continued operation can we have useable quantities of equipment available over time.

In order to make use of this model, the crucial parameter to be estimated is the effectiveness factor, a. Once a value is assigned to this factor and also to the budget equation

[3]Strictly speaking, this correction factor is applicable only to cases in which the retirement age y is an integral multiple of the development interval, z. Furthermore, this correction factor is intended to be applicable only at the beginning and end of each production run, i.e., at the low point of the effectiveness cycle.

parameters, it becomes a comparatively simple matter to work out both the optimal equipment replacement interval and the optimal spacing between successive new models. At a subsequent point, curves will be presented that indicate the optimal magnitude of both these intervals over a range of values for the effectiveness factor, a.

THE MATHEMATICAL MODEL

Some definitions:

Decision Variables:

- x = annual rate of initial expenditures on new million-dollar military units ($ millions/year)
- y = retirement age of equipment (years)
- z = time interval between successive new types (years)

Components of the budget equation:

- B = annual budgetary limitation ($ millions/year)
- cxy = annual operating expenses for a force with initial expenditures occurring at a steady rate x and with equipment retired after y years of use ($ millions/year)
- d/z = annual rate of development expenditures ($ millions/year)

a^t = comparative effectiveness of a military unit of age t

$\int_{t=o}^{y} a^t \, dt$ = equivalent military worth of brand-new equipment represented by continuously procuring at the annual rate of one million dollars' worth of the best state-of-the-art equipment available, and by continuously retiring equipment at the end of y years (area ABCDFGH on Figure 1)

$\frac{z\,a^z}{\int_{t=o}^{z} a^t \, dt}$ = fraction of military effectiveness remaining after use of a "staircase" approximation to the area ABCDFGH; on Figure 1 this fraction equals ABDE/ABCDE

$E(x, y, z)$ = equivalent military worth in terms of brand-new equipment represented by a force with initial procurement expenditures occurring at a steady rate x, with equipment retired after y years of use, and with new types introduced every z years (unit: $ millions' worth of brand-new equipment)

The military effectiveness function ($E(x, y, z)$) measures the current military worth of the equipment in each age bracket, summed up over all vintages represented in the inventory, and after correcting for the "staircase" approximation:

617404 O - 61 - 3

$$(1) \qquad E(x, y, z) = \left[\frac{z\, a^z}{\int_{t=o}^{z} a^t\, dt}\right]\left[\int_{t=o}^{y} xa^t\, dt\right],$$

$$(2) \qquad E(x, y, z) = x\,[f(z)]\,[1 - a^y],$$

where $f(z) = \dfrac{za^z}{1 - a^z}$.

The optimization problem now consists of choosing $\underline{x}$, $\underline{y}$, and $\underline{z}$ so as to maximize this measure of effectiveness, subject to the budgetary constraint on the total of expenditures for procurement, operation, and development:

$$(3) \qquad x + cxy + \frac{d}{z} = B.$$

This constrained maximum problem may also be transformed into an unconstrained one, where the following function is to be optimized simultaneously with respect to x, y, z, and the Lagrangean multiplier, λ:

$$\phi(x, y, z, \lambda) = E(x, y, z) + \lambda\left(x + cxy + \frac{d}{z} - B\right).$$

At an interior optimum:

$$(4) \qquad \frac{\partial\, \phi(x, y, z, \lambda)}{\partial x} = [f(z)]\,[1 - a^y] + \lambda[1 + cy] = 0,$$

$$(5) \qquad \frac{\partial\, \phi(x, y, z, \lambda)}{\partial y} = x[f(z)]\,[-a^y \log a] + \lambda cx = 0,$$

$$(6) \qquad \frac{\partial\, \phi(x, y, z, \lambda)}{\partial z} = x[1 - a^y]\,[f'(z)] - \frac{\lambda d}{z^2} = 0,$$

$$(7) \qquad \frac{\partial\, \phi(x, y, z, \lambda)}{\partial \lambda} = x + cxy + \frac{d}{z} - B = 0.$$

By combining (4) and (5), one may solve for the unknown optimal value y^*. In implicit form, one obtains:

$$(8) \qquad \left[\frac{1}{\log a}\right]\left[1 - a^{-y^*}\right] = \frac{1}{c} + y^*.$$

This means that we may sub-optimize for y^*, the optimal value for the retirement age of equipment. This variable is dependent upon only two numerical parameters: $\underline{c}$, the annual

operating cost factor, and a, the effectiveness decline factor. Neither the magnitude of development costs, d, nor the absolute size of the budget, B, influences y^*—although these obviously do affect the optimal values x^* and z^*.

Through conditions (5), (6), and (7), the reader may now verify certain equalities:

By (5):
$$\lambda = \left[\frac{1}{c}\right]\left[f(z)\right]\left[a^{y^*} \log a\right],$$

By (7):

(9)
$$x = \frac{B(z - d/B)}{z(1 + cy^*)},$$

And by (6):

(10)
$$\left[\frac{(z - d/B)}{z(1 + cy^*)}\right]\left[1 - a^{y^*}\right]\left[f'(z)\right] = \frac{[d/B]\,[f(z)]\,[a^{y^*} \log a]}{cz^2}.$$

According to (10), the optimal value of z does not depend upon the absolute values of all the parameters: a, B, c, and d. Rather, this unknown depends upon the absolute values of a and c and upon the ratio d/B. Hence, if our interest lies in optimization with respect to the development spacing interval z, we are at liberty to normalize by setting the budget scale, B, at an arbitrary level, say a billion dollars a year. The resulting value of z^* will remain optimal for any other budget level—just as long as the ratio d/B is kept unchanged.

NUMERICAL ANALYSIS

In order to determine y^* numerically from Eq. (8), the most convenient method appears to be: First select a series of trial values for y^*. Then separately calculate the magnitude of the left- and the right-hand side of Eq. (8). And finally use graphical interpolation to select the trial value of y^* for which the right-hand side equals the left. This was the method employed in constructing Figure 2.

In view of the clumsiness of Eq. (10), the remaining numerical analysis was performed as follows: (i) Assign numerical values to the parameters a, c, and d/B. (Throughout, the budget level is maintained at its normalized level B = \$1,000 millions per year.) (ii) Read off y^* from Figure 2. (iii) Select a trial value for z, and then use Eq. (9) to solve for the corresponding trial value of x. (iv) Evaluate the function $E(x, y^*, z)$ at these trial values. (v) Vary the numerical value of z and adjust x simultaneously so as to find the approximate maximum of $E(x, y^*, z)$.

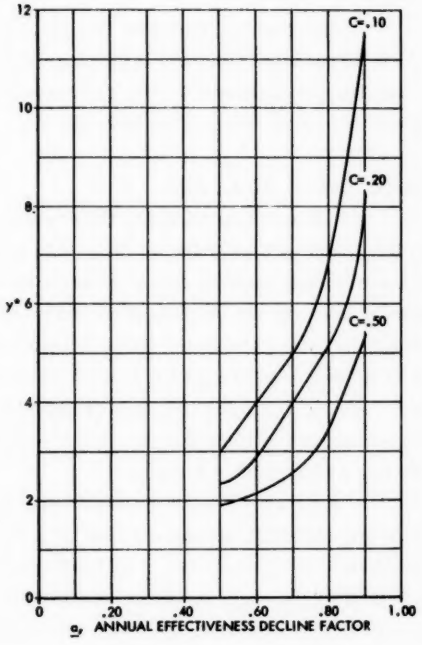

Figure 2 - y^*, optimal retirement age (years)

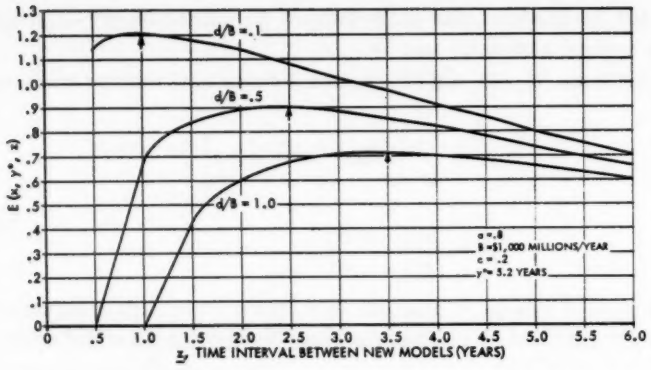

Figure 3 - Relative current effectiveness--initial cost of equivalent amount of brand-new equipment (unit: $ billion)

Figure 3 contains an example in which this procedure was employed. Throughout, we hold constant the effectiveness factor, a, the budget level, B, and the cost parameter, c. (a = .8; B = 1,000; c = .2; and from Figure 3, y* = 5.2 yr.) Three different cases are shown in Figure 3. These correspond to high, low, and medium values for the ratio d/B. As this d/B ratio is increased—i.e., as the relative cost of each individual development effort gets larger—the optimal spacing between successive models becomes stretched out from 1.0 to 2.5 to 3.5 yr.

Note that the function E(x, y*, z) generally remains quite flat within the neighborhood of ±6-12 months about the optimum. Hence little loss in military effectiveness would result from the assumption that the retirement age be an integral multiple of the time interval between new models. (See footnote 3.)

SENSITIVITY ANALYSIS

In attempting to apply this model, it is well to recall that there is an inherent difference in the kinds of precision which can be associated with the parameters a, c, and d/B. The economic cost factors can ordinarily be estimated in a considerably more objective fashion than a, the rate of decline of military usefulness. Figure 3 enables the military planner to insert his own judgments about the magnitude of the decline rate, and to read out how the optimal retirement age y* is then affected. If, for example, we maintain c constant at .2 and decrease a from .9 to .6, this change has a substantial effect upon the optimal retirement age—a decrease from 8.2 to 2.9 yr. (Compare this with the days of Trafalgar and Nelson's 40-yr-old flagship, "Victory.")

Similarly, Figure 4 enables the planner to insert his own judgments about the effectiveness decline rate, and to read off z*, the optimal development spacing. On this figure, the operating cost parameter c is held constant at the representative level of .2, and the results are plotted for high, medium, and low values of d/B. Note that the more rapid the progress in military technology, the lower the value of a, and the closer becomes the optimal spacing. With d/B = .5, for example, and with a reduced from .9 to .6, the interval shrinks from 3.5 to 1.5 yr.

One nontrivial implication of this model is that "the rich get richer." Consider the parable of two countries—Davidland and Goliathland. Both are acquainted with the same technology and with the same effectiveness parameter, a. Costs are the same in both countries. Hence the parameters c and d are identical, and the optimal retirement age is the same in both places. The one difference is that Goliathland has a military budget twice that of Davidland. The fact that its d/B ratio is half that of its enemy means that Goliathland can afford to keep bringing out new types of equipment at far more frequent intervals, and that it can thereby squeeze out a substantially larger share of the military benefits conferred by an exponentially growing level of military technology. Not a rosy prospect for Davidland!

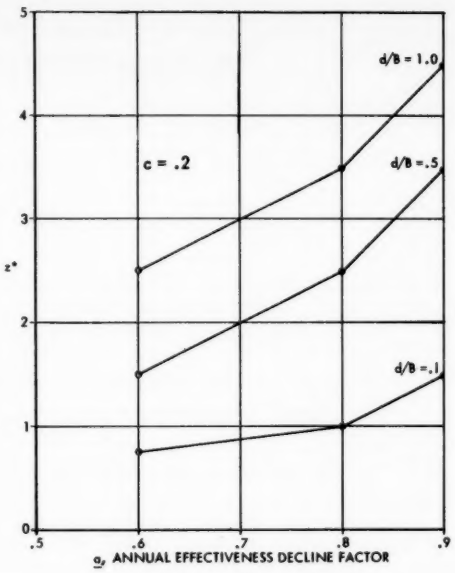

Figure 4 - z*, optimal development interval (years)

* * *

ON THE ESTIMATION OF SCALE PARAMETERS*†

Lionel Weiss

Cornell University

If $Y_1 \leq \ldots \leq Y_n$ are ordered observations from a population with cumulative distribution function $G\left(\frac{x - B}{C}\right)$, probability density function $(1/C)g\left(\frac{x - B}{C}\right)$, where B and C are unknown parameters with $C > 0$, and the function G(x) is known, it is shown that under mild restrictions on G(x), $\frac{n+1}{n-1} \sum_{j=1}^{n-1} g[G^{-1}(j/n)]\,(Y_{j+1} - Y_j)$ is a consistent estimate of C. In certain important cases, this estimate has a structure similar to that of estimates known to be optimal.

INTRODUCTION

One of the most common problems in statistics is the problem of estimating a scale parameter. In many cases, we are willing to assume that the underlying probability distribution of the random variable of interest has a certain form because of certain theoretical considerations, but these theoretical considerations may not be of any help in specifying the parameters of the distribution. To take a common case, it may be safe to assume that the total demand for a certain item during a given time period has a normal distribution, because the total demand is the sum of many independent individual demands, and the central limit theorem tells us that in such a case the total demand has approximately a normal distribution. But this reasoning still leaves unspecified the parameters of the normal distribution, which will presumably have to be estimated from observed data. One of these parameters is the standard deviation, which is a scale parameter. To take another common case, we may be willing to assume that the total length of life of a certain piece of equipment has an exponential distribution, because it is felt that the failure rate is constant. However, the actual value of the failure rate may not be known, in which case it would have to be estimated from observed data. The failure rate is a scale parameter in the exponential distribution of length of life.

Of course, the problem of estimating the scale parameters in the normal distribution and the exponential distribution has been completely solved, in the sense that it is known explicitly which estimate is optimal. But recently there has been a tendency to use certain probability distributions containing a scale parameter, where the optimal estimate of the scale parameter is not known. A familiar case of this sort is the recent tendency to assume that the

*Manuscript received August 3, 1960.
†Research supported by National Science Foundation Grant No. NSF-G11321.

total length of life of a given piece of equipment has a Weibull distribution. The Weibull distribution is a generalization of the exponential distribution, and therefore allows more flexibility in fitting to observed data. But there are serious problems in trying to estimate the parameters of a Weibull distribution.

Various general techniques for estimating scale parameters are in use: the maximum likelihood estimate, the best invariant estimate [1], and the method of using a linear combination of order statistics. In some important cases, the maximum likelihood estimate and the best invariant estimate cannot be found in closed form. In most cases, finding the best linear combination of order statistics requires the use of large-scale computing equipment [2].

In this paper, a heuristic development leads to an explicit linear combination of sample spacings as a proposed estimate of the scale parameter. It is shown that in the important cases of the exponential and normal distributions, the optimal estimate has a structure like that of the proposed estimate. Then it is shown that under mild restrictions on the distribution, the proposed estimate is consistent. The estimate is applied to the Weibull distribution. The relation of the estimate to others proposed recently is discussed.

THE HEURISTIC DEVELOPMENT OF THE ESTIMATE

$X_1, X_2, \ldots, X_n$ are independent random variables, each with the same continuous cumulative distribution function $F(x)$. $F(x)$ is not known exactly, but it is known that $F(x) = G\left(\frac{x - B}{C}\right)$, where B, C are unknown constants with $C > 0$, and the cumulative distribution function $G(x)$ is known. The derivatives of $F(x)$, $G(x)$ are denoted by $f(x)$, $g(x)$ respectively. In the usual terminology, B is a location parameter and C is a scale parameter. The problem is to find a point estimate for C.

Denote the ordered values of $X_1, \ldots, X_n$ by $Y_1, \ldots, Y_n$, where $Y_1 \leq Y_2 \leq \ldots \leq Y_n$. Denote $Y_{j+1} - Y_j$ by W_j for $j = 1, \ldots, n - 1$, and denote $g[G^{-1}(j/n)]W_j$ by T_j.

To the author's knowledge, every estimate of C that has been proposed can be written in the form $M(T_1, \ldots, T_{n-1})$, where $M(T_1, \ldots, T_{n-1})$ is a homogeneous function of order one. That is, for any positive λ, $M(\lambda T_1, \ldots, \lambda T_{n-1}) = \lambda M(T_1, \ldots, T_{n-1})$. Then if $M(T_1, \ldots, T_{n-1})$ is differentiable with respect to each of its arguments, Euler's formula gives $M(T_1, \ldots, T_{n-1}) = T_1 \partial M/\partial T_1 + \ldots + T_{n-1} \partial M/\partial T_{n-1}$, where $\partial M/\partial T_j$ is homogeneous of order zero for $j = 1, \ldots, n - 1$. Since a function which is homogeneous of order zero in $T_1, \ldots, T_{n-1}$ clearly contains little if any information about the scale parameter C, intuitively we may expect $\partial M/\partial T_j$ to converge stochastically to a constant as n increases.

Suppose we assume that $\partial M/\partial T_j$ depends only on j and n for each j, and investigate what a reasonable value for the quantity might be. We note that the random variables $F(Y_1), \ldots, F(Y_n)$ are distributed as the n ordered observations from a uniform distribution over $(0, 1)$. We have $F(Y_{j+1}) - F(Y_j) = f(\theta_j)(Y_{j+1} - Y_j)$, where θ_j is between Y_j and Y_{j+1}. When n is large, the Glivenko-Cantelli theorem states that with high probability, Y_j is close to $F^{-1}(j/n)$, so that θ_j is close to $F^{-1}(j/n)$ with high probability. Therefore $F(Y_{j+1}) - F(Y_j)$ can be approximated by $f[F^{-1}(j/n)](Y_{j+1} - Y_j) = (1/C)\, g[G^{-1}(j/n)](Y_{j+1} - Y_j) = T_j/C$, with high probability for large n. Thus $M(T_1, \ldots, T_{n-1})$ can be approximated by

$$C \sum_{j=1}^{n-1} \left(\partial M/\partial T_j\right)\left[F\left(Y_{j+1}\right) - F\left(Y_j\right)\right].$$

Assuming that $\partial M/\partial T_j$ depends only on j and n, we can ask which choice of value for $\partial M/\partial T_j$ will make the expression

$$C \sum_{j=1}^{n-1} \left(\partial M/\partial T_j\right)\left[F\left(Y_{j+1}\right) - F\left(Y_j\right)\right]$$

a good estimate of C. Alternatively, we may ask which choice of value for $\partial M/\partial T_j$ will cause

$$\sum_{j=1}^{n-1} \left(\partial M/\partial T_j\right)\left[F\left(Y_{j+1}\right) - F\left(Y_j\right)\right]$$

to be a good estimate of unity. Because the joint distribution of the n-1 quantities $F(Y_2) - F(Y_1)$, $F(Y_3) - F(Y_2)$, ..., $F(Y_n) - F(Y_{n-1})$ is symmetric in these quantities, and $F(Y_1)$, $F(Y_n)$ converge stochastically to zero, unity respectively as n increases, the best choice for $\partial M/\partial T_j$ would be a value K_n depending only on n, where K_n approaches unity as n increases. In particular, setting K_n equal to $(n + 1)/(n - 1)$ gives an unbiased estimate of unity. Then $M(T_1, \ldots, T_{n-1})$ becomes

$$\frac{n+1}{n-1} \sum_{j=1}^{n-1} T_j = \frac{n+1}{n-1} \sum_{j=1}^{n-1} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right).$$

This is the estimate of the scale parameter C that will be investigated in this paper.

Next, we investigate the relation of our proposed estimate to estimates which are known to be optimal. The first case we discuss is the exponential distribution, where

$$F(x) = 1 - \exp\left\{-\left(\frac{x - B}{C}\right)\right\}$$

for $x \geq B$. Thus $G(x) = 1 - e^{-x}$ for $x \geq 0$, and $g(x) = e^{-x}$ for $x > 0$. It is easily verified that $g[G^{-1}(v)] = 1 - v$ for all v in the open interval (0, 1). Thus, our proposed estimate in this case is

$$\frac{n+1}{n-1} \sum_{j=1}^{n-1} (1 - j/n) W_j.$$

The optimal estimate in this case is known to be

$$\frac{1}{n-1}\sum_{j=2}^{n}\left(Y_j - Y_1\right).$$

Since $Y_j - Y_1 = W_1 + \ldots + W_{j-1}$ for $j = 2, \ldots, n$, the optimal estimate is

$$\frac{1}{n-1}\sum_{j=2}^{n}\left(W_1 + \ldots + W_{j-1}\right) = \frac{1}{n-1}\sum_{j=1}^{n-1}(n-j)\,W_j = \sum_{j=1}^{n-1}[1-(j-1)/(n-1)]\,W_j .$$

Thus the coefficient of T_j in the optimal estimate is asymptotically equivalent to the coefficient in the proposed estimate.

The second case we discuss is the normal distribution, where

$$f(x) = (1/C\sqrt{2\pi})\exp\left\{-\frac{1}{2}\left(\frac{x-B}{C}\right)^2\right\}.$$

The optimal estimate in this case is known to be

$$S_n = K_n\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]^{\frac{1}{2}},$$

where $\overline{Y} = (Y_1 + \ldots + Y_n)/n$, and K_n is a constant which approaches unity as n increases. Since

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=2}^{n}\left(Y_i - Y_1\right)^2 - n\left(Y_1 - \overline{Y}\right)^2,$$

and since $Y_i - Y_1 = W_1 + \ldots + W_{i-1}$ for $i = 2, \ldots, n$, while $Y_1 - \overline{Y} = (-1/n)[(n-1)W_1 + \ldots + W_{n-1}]$, we find that

$$\partial\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2/\partial W_j = 2\sum_{i=j+1}^{n}\left(Y_i - Y_1\right) + 2(n-j)\left(Y_1 - \overline{Y}\right) = 2\sum_{i=j+1}^{n}\left(Y_i - \overline{Y}\right),$$

and we have that

$$\partial S_n/\partial W_j = K_n^2 S_n^{-1} (1/n) \sum_{i=j+1}^{n} \left(Y_i - \bar{Y}\right).$$

Let r be a fixed value in the open interval (0, 1). As n increases, K_n approaches unity, S_n^{-1} converges stochastically to 1/C, $\bar{Y}$ converges stochastically to B, and

$$(1/n) \sum_{i=rn+1}^{n} \left(Y_i - \bar{Y}\right)$$

converges stochastically to

$$(1/C\sqrt{2\pi}) \int_{F^{-1}(r)}^{\infty} (y - B) \exp\left\{-\frac{1}{2}\left(\frac{y - B}{C}\right)^2\right\} dy = (C/\sqrt{2\pi}) \exp\left\{-\frac{1}{2}\left[\frac{F^{-1}(r) - B}{C}\right]^2\right\} =$$

$$C\, g\left[\frac{F^{-1}(r) - B}{C}\right] = C\, g\left[G^{-1}(r)\right].$$

Therefore, $\partial S_n/\partial W_{rn}$ converges stochastically to $g\left[G^{-1}(r)\right]$ as n increases. This means that $\partial S_n/\partial T_{rn}$ converges stochastically to unity as n increases. Once again, the coefficient of T_j in the optimal estimate is asymptotically equivalent to the coefficient in the proposed estimate.

THE CONSISTENCY OF THE PROPOSED ESTIMATE

The discussion in the preceding section was heuristic. In this section, we prove that the estimate proposed in the preceding section is a consistent estimate of C, if the cumulative distribution function G(x) satisfies the following four conditions:

(1) There exists a finite value K such that $g(x) \leq K$ for all x, g(x) has a finite number of discontinuities, and $\int_{-\infty}^{\infty} x\, g(x)\, dx$ is finite.

(2) There exists a positive value D' such that for any value D in the open interval (0, D'), the set of points $\{x: g(x) \geq D\}$ is an interval.

(3) If there is no finite value y such that G(y) = 1, then there exists a positive constant L and a value v in the open interval (0, 1), such that $dg[G^{-1}(r)]/dr < -L$, for all r in the open interval (v, 1).

(4) If there is no finite value y such that $G(y) = 0$, then there exists a positive constant M and a value u in the open interval (0, 1), such that $dg[G^{-1}(r)]/dr > M$ for all r in the open interval (0, u).

It will be seen that the consistency of our estimate can be proved under somewhat weaker assumptions. However, the assumptions just listed hold in most cases of interest, and make the proof of consistency simpler. It should be noted that Condition (2) is satisfied by every unimodal g(x). Conditions (3) and (4) are usually simple to verify by means of the fact that $dg[G^{-1}(r)]/dr = d \log g[G^{-1}(r)]/dG^{-1}(r)$. Thus, if there is no finite value y such that $G(y) = 1$, then Condition (3) is equivalent to stating that there is a finite value q such that $d \log g(y)/dy < -L$ for all $y > q$. In the exponential case, $\log g(y) = -y$, $d \log g(y)/dy = -1$, so that Condition (3) is satisfied. In the normal case, $\log g(y) = -\frac{1}{2} \log 2\pi - \frac{1}{2} y^2$, $d \log g(y)/dy = -y$, and Condition (3) is satisfied. Similarly, if there is no finite value y such that $G(y) = 0$, Condition (4) is equivalent to stating that there is a finite value p such that $d \log g(y)/dy > M$ for all $y < p$. In the case of a normal distribution, Condition (4) is satisfied.

Now we turn to the proof of consistency, assuming G(x) satisfies Conditions (1)-(4). Throughout the discussion, we replace $(n + 1)/(n - 1)$ by unity, since we study what happens as n increases. For any value D in the open interval $(0, D'/C)$, let A_D and A'_D denote the lower and upper endpoints, respectively, of the interval $\left\{x: (1/C)\, g\left(\frac{x - B}{C}\right) \geq D\right\}$. Denote the quantity $G\left(\frac{A_D - B}{C}\right)$ by u_D, and the quantity $G\left(\frac{A'_D - B}{C}\right)$ by v_D. It follows directly from Theorem 2.2 of [3], that

$$\sum_{nu_D < j < nv_D} g\left[G^{-1}(j/n)\right] \left(Y_{j+1} - Y_j\right)$$

converges stochastically to

$$\int_{u_D}^{v_D} \frac{g[G^{-1}(x)]}{f[F^{-1}(x)]}\, dx = C\left(v_D - u_D\right),$$

as n increases.

Next, we investigate the quantity

$$\sum_{j \geq nv_D} g\left[G^{-1}(j/n)\right] \left(Y_{j+1} - Y_j\right).$$

First, we take the case where there is a finite value y with $G(y) = 1$. Let β denote $\inf\left\{z: G\left(\frac{z - B}{C}\right) = 1\right\}$. Then, Y_n converges stochastically to β as n increases. Also,

$$\sum_{j \geq nv_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) \leq K \sum_{j \geq nv_D} \left(Y_{j+1} - Y_j\right) = K\left(Y_n - Y_{[nv_D]}\right) \leq K\left(\beta - Y_{[nv_D]}\right),$$

where $[nv_D]$ denotes the smallest integer which is not less than nv_D. $Y_{[nv_D]}$ converges stochastically to A'_D as n increases, and A'_D approaches β as D approaches zero. Thus, given any positive ε, we can find a value D* in the open interval $(0, D'/C)$ such that

$$P\left[\sum_{j \geq nv_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) \leq \varepsilon\right]$$

approaches unity as n increases, for each D in the open interval $(0, D^*)$.

Now we investigate the quantity

$$\sum_{j \geq nv_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right)$$

when $G(y) < 1$ for any finite value of y. We take D small enough so that $d\, g[G^{-1}(r)]/dr < -L$ for all $r \geq v_D$. For $r \geq v_D$, we define a function $h[G^{-1}(r)]$ as follows: $h[G^{-1}(v_D)] = g[G^{-1}(v_D)]$, and $d\, h[G^{-1}(r)]/dr = -L$ for all $r \geq v_D$. Then it is clear that $h[G^{-1}(r)] \geq g[G^{-1}(r)]$ for all $r \geq v_D$, and therefore

$$\sum_{j \geq nv_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) \leq \sum_{j \geq nv_D} h\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) =$$

$$-Y_{[nv_D]}\, h\left[G^{-1}([nv_D]/n)\right] + \sum_{j > [nv_D]} \left\{h\left[G^{-1}((j-1)/n)\right] - h\left[G^{-1}(j/n)\right]\right\} Y_j =$$

$$-Y_{[nv_D]}\, h\left[G^{-1}([nv_D]/n)\right] + (L/n) \sum_{j > [nv_D]} Y_j .$$

But as n increases, $h[G^{-1}([nv_D]/n)]$ approaches

$$g\left[G^{-1}(v_D)\right] = g\left(\frac{A'_D - B}{C}\right),$$

$Y_{[nv_D]}$ converges stochastically to A'_D, and

$$(L/n) \sum_{j > [nv_D]} Y_j$$

converges stochastically to

$$(L/C) \int_{A'_D}^{\infty} y\, g\left(\frac{y - B}{C}\right) dy .$$

Therefore

$$\sum_{j \geq nv_D} h\left[G^{-1}(j/n)\right] \left(Y_{j+1} - Y_j\right)$$

converges stochastically to

$$-A'_D\, g\left(\frac{A'_D - B}{C}\right) + (L/C) \int_{A'_D}^{\infty} y\, g\left(\frac{y - B}{C}\right) dy .$$

Since by Condition (1),

$$\int_{-\infty}^{\infty} y\, g\left(\frac{y - B}{C}\right) dy$$

is finite, and since A'_D approaches infinity as D approaches zero, it is clear that

$$-A'_D\, g\left(\frac{A'_D - B}{C}\right) + (L/C) \int_{A'_D}^{\infty} y\, g\left(\frac{y - B}{C}\right) dy$$

approaches zero as D approaches zero. Thus, given any positive ε, we can find a value D^* in the open interval $(0, D'/C)$, such that

$$P\left[\sum_{j \geq nv_D} g\left[G^{-1}(j/n)\right] \left(Y_{j+1} - Y_j\right) \leq \varepsilon\right]$$

approaches unity as n increases, for each D in the open interval $(0, D^*)$.

By exactly the same sort of reasoning, we can show that given any positive ε, we can find a value D^* in the open interval $(0, D'/C)$, such that

$$P\left[\sum_{j \le nu_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) \le \varepsilon\right]$$

approaches unity as n increases, for each D in the open interval $(0, D^*)$.

From the discussion just completed, we see that we can write

$$\sum_{j=1}^{n-1} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) \quad \text{as} \quad \sum_{nu_D < j < nv_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right) + S(D,n),$$

where $S(D,n)$ is a random variable with the property that for any given positive ε, we can find a value D^* such that $P[S(D,n) < \varepsilon]$ approaches unity as n increases, for each D in the open interval $(0, D^*)$. Since

$$\sum_{nu_D < j < nv_D} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right)$$

converges stochastically to $C(v_D - u_D)$ as n increases, and since $v_D - u_D$ approaches unity as D approaches zero, it is clear that

$$\sum_{j=1}^{n-1} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right)$$

converges stochastically to C as n increases. This implies that

$$\frac{n+1}{n-1} \sum_{j=1}^{n-1} g\left[G^{-1}(j/n)\right]\left(Y_{j+1} - Y_j\right)$$

is a consistent estimate of C.

APPLICATION TO THE WEIBULL DISTRIBUTION

If the random variable Z has a Weibull distribution, then $P(Z \le z) = 1 - \exp\{-(z/\theta)^{1/C}\}$, for positive z. Here θ and C are positive parameters. θ is a scale parameter, C a "shape" parameter. Often we want to estimate C, but standard techniques such as the maximum likelihood method lead to intractable equations. Defining the random variable X as log Z, and denoting θ by e^B, we find that $P(X \le x) = 1 - \exp\{-\exp[(x - B)/C]\}$. Thus B, C are location and scale parameters, respectively, in the distribution of X. For the remainder of this

section, we assume that we are working in terms of the random variable X. The problem is to estimate the parameter C, both B and C being unknown.

Defining the cumulative distribution function G(x) as $1 - \exp(-e^x)$, we have $g(x) = e^x \exp(-e^x)$, $\log g(x) = x - e^x$, $d \log g(x)/dx = 1 - e^x$. It is easily verified that G(x) satisfies conditions (1)-(4) of the preceding section. Letting $Y_1 \le \ldots \le Y_n$ denote the ordered values of $X_1, \ldots, X_n$, where $X_1, \ldots, X_n$ are independent random variables, each with the cumulative distribution function

$$G\left(\frac{x - B}{C}\right),$$

we know that

$$\frac{n+1}{n-1} \sum_{j=1}^{n-1} g\left[G^{-1}(j/n)\right] \left(Y_{j+1} - Y_j\right)$$

is a consistent estimate of C. Here $G^{-1}(r) = \log \log (1/(1-r))$, and $g[G^{-1}(r)] = -(1-r)\log(1-r)$. Therefore

$$-\frac{n+1}{n-1} \sum_{j=1}^{n-1} [(1 - j/n) \log (1 - j/n)] \left(Y_{j+1} - Y_j\right)$$

is a consistent estimate of C.

RELATION TO OTHER WORK

In [2], for the problem of estimating the scale parameter of a normal distribution, that particular linear combination of $Y_1, \ldots, Y_n$ is computed which gives a minimum variance unbiased estimate among all linear combinations of $Y_1, \ldots, Y_n$, for values of n up to 15. Table 1 compares the coefficients of $Y_1, \ldots, Y_n$ given by the estimate proposed in this article to the best coefficients from [2], for the case n = 15. The two sets of coefficients seem satisfactorily close, particularly since a sample size of 15 is not exceptionally large.

Jung and Blom, in [4] and [5], have developed asymptotic approximations to the mean and variance of linear combinations of $Y_1, \ldots, Y_n$, and using these asymptotic approximations, have found the linear combination of $Y_1, \ldots, Y_n$ which is an asymptotically unbiased estimate of $k_1B + k_2C$ with asymptotically minimum variance, where k_1 and k_2 are specified constants, and B, C are the unknown location and scale parameters, respectively. The coefficients given by Jung and Blom are not given as explicitly as in this paper. In those examples worked out explicitly in [5], the coefficients are asymptotically equivalent to the coefficients given in this paper. Thus on page 139 of [5], for the problem of estimating the unknown scale parameter of a normal distribution, the asymptotically best coefficient of Y_j is given as $(1/A)G^{-1}(j/(n+1))$, where A is defined as

$$\sum_{i=1}^{n} h_i \, G^{-1}(i/(n+1)) ,$$

TABLE 1

Comparison of the Coefficients of $Y_1, \ldots, Y_n$ Given by the Estimate Proposed in this Article to the Best Coefficients from [2], for the Case n = 15

j	Coefficient of Y_j in Proposed Estimate	Best Coefficient of Y_j
1	-0.1478	-0.1444
2	-0.0982	-0.0927
3	-0.0741	-0.0699
4	-0.0555	-0.0526
5	-0.0400	-0.0379
6	-0.0259	-0.0247
7	-0.0128	-0.0122
8	0.0000	0.0000
9	0.0128	0.0122
10	0.0259	0.0247
11	0.0400	0.0379
12	0.0555	0.0526
13	0.0741	0.0699
14	0.0982	0.0927
15	0.1478	0.1444

and h_i is the expected value of the i^{th} order statistic out of n observations from a standard normal distribution. The coefficient of Y_j in the estimate proposed in this paper is

$$\frac{n+1}{n-1}\left\{-g\left[G^{-1}(j/n)\right] + g\left[G^{-1}((j-1)/n)\right]\right\},$$

which is approximately equal to $-(1/n)d\,g[G^{-1}(r)]/dr$ evaluated at $r = j/n$. For the special case when G(x) is the standard normal cumulative distribution function, this last expression is equal to $\frac{G^{-1}(j/n)}{n}$. Thus, the coefficient of Y_j in the proposed estimate when the scale parameter of a normal distribution is being estimated is approximately $\frac{G^{-1}(j/n)}{n}$, and clearly this must be asymptotically equivalent to $(1/A)G^{-1}(j/(n+1))$, since both estimates are asymptotically unbiased.

Jung and Blom prove that under certain regularity conditions on G(x) the estimates they construct are asymptotically efficient among all possible estimates.

617404 O - 61 - 4

REFERENCES

[1] E. J. G. Pitman, "The Estimation of the Location and Scale Parameters of a Continuous Population of any Given Form," Biometrika 30: p. 391 (1939).

[2] A. E. Sarhan and B. G. Greenberg, "Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples. Part I," Ann. Math. Stat. 27: p. 427 (1956).

[3] Lionel Weiss, "Tests of Fit in the Presence of Nuisance Location and Scale Parameters," Ann. Math. Stat. 28: p. 1016 (1957).

[4] Jan Jung, "On Linear Estimates Defined by a Continuous Weight Function," Arkiv För Matematik 3: p. 199 (1956).

[5] Gunnar Blom, "Statistical Estimates and Transformed Beta-Variables," John Wiley and Sons, Inc., New York (1958).

* * *

A TRANSPORTATION ALGORITHM AND CODE*

Merrill M. Flood

The University of Michigan

A FORTRAN II transportation code, using Kuhn's Hungarian Method, was reported upon at the RAND Symposium on Mathematical Programming in March 1959. The algorithm was based upon a proof of the König-Egervary theorem, presented by the present author at the 1958 Symposium on Combinatorial Problems sponsored by the American Mathematical Society. The code was entirely rewritten (in FORTRAN II), during the 1959 IBM Summer Institute on Combinatorial Problems, and several problems were run at the IBM Research Center to obtain data regarding computing times and frequencies of various internal loops. A CLOCK Subroutine yields readings, in hundredths of minutes, for each time through selected portions of the computing run. The version reported upon at RAND solved a 29 x 116 pseudorandom transportation problem in 8.01 min, as compared with 3.17 min using the fastest competing code (SHARE 464) then available. The present version solved this same problem in 2.87 min, and another 29 x 116 problem in 1.89 min. This paper presents the algorithm and reports upon computing experience.

INTRODUCTION

The transportation problem and the assignment problem are two alternative names for the Hitchcock distribution problem[1] [1]. This problem has received considerable attention in recent years because of its importance in a wide variety of practical situations, and especially for industrial operations research applications.

We refer the interested reader to other sources [2] for a discussion of practical applications, and for further references to the many techniques that have been offered for solving this problem.

Solution techniques seem to fall into two broad classes. One class consists of variants of the Simplex Method of Dantzig [3], including the earliest work by Hitchcock [4] and Kantorovich [5]; also including some previous work by the present author [6]. The other class consists of variants of the Hungarian Method of Kuhn [7], including the present paper. Generally speaking, the Hungarian Method seems to be superior to the Simplex Method, for this special case of the general linear programming problem, but efforts to extend the idea of the Hungarian Method to the general linear programming problem have not produced computational techniques superior to the Simplex Method for the more general case.

*Manuscript submitted January 3, 1961.

[1] Some authors have distinguished between the assignment and transportation problems, and this may sometimes be convenient. However, since each problem is algebraically a special case of the other, and since any computational method that will solve one problem also solves the other, the problems are mathematically identical [1]. The present paper makes no distinction between the two problems.

The Hungarian Method, and the Hitchcock distribution problem, seem important enough to merit further development; the present paper discusses a new variant that has proved to be effective computationally.[2]

PROBLEM

The assignment problem is stated as follows:

Devise an efficient computational algorithm for permuting the rows of a square matrix of non-negative integers so as to minimize the sum of the elements on the main diagonal.

If the rows (or columns) of the matrix are not distinct then this special case is known as the transportation problem, and is exactly the product distribution problem of Hitchcock [4] when stated in this form.

The problem may be rewritten in linear programming form as follows: Find nonnegative integers x_{ij}, subject to the restrictions

$$\sum_{j=1}^{n} x_{ij} = r_i \quad \text{and} \quad \sum_{i=1}^{m} x_{ij} = c_j ,$$

that minimize the quantity

$$\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} x_{ij} ;$$

where r_i, c_j, d_{ij}, m, and n are given positive integers. Under this formulation, the problem is usually called an assignment problem if $r_i = c_j = 1$, otherwise it is usually called a transportation problem.

HUNGARIAN METHOD

It is evident that the solution to the assignment problem is not changed by adding a constant to each element of a "line" of the matrix, where a line is either a row or a column. It is also evident that if the addition of such constants yields a matrix of nonnegative elements, but such that the main diagonal is zero after some permutation of the rows, then this permutation is the solution we seek. The Hungarian Method proceeds in this fashion, always keeping the matrix nonnegative while increasing the number of null elements that can be brought to the main diagonal by a row permutation.

The mathematical basis for the Hungarian Method is provided by a constructive proof of the following theorem of Konig [8] and Egervary [9]:

[2] Acknowledgment. The author is very grateful to the IBM Research Center for supporting work on the computer program for this algorithm, and for the use of the IBM 704 at the Center in testing the computational quality of the program. This work was done largely during two summers at the Center (while the author served as a Consultant to IBM), including the period of the Combinatorial Problems Institute during July of 1959. The author is especially indebted to V. V. Van Ness of the Center for programming advice and for generous help in running test problems on the Center's IBM 704.

THEOREM (KÖNIG-EGERVARY): If D is a square array of two kinds of marks, say zeros and plusses, and if:

1. x is the maximum number of zeros that can be found in the array such that no two of them are in the same line, and

2. y is the minimum number of lines that can be found such that each zero of the array is contained in one of them, then x = y.

The present author has shown elsewhere [10] how the rows and columns of a (0, +) matrix can be permuted into the following "standard" form.

	1	2	3	4	- - - - -	m-2	m-1	m	m+1	m+2	m+3	m+4
1	0	c	+	+	- - - - -	+	+	+	+	+	+	+
2		0	c	+	- - - - -	+	+	+	+	+	+	+
3			0	c		+	+	+	+	+	+	+
4				0		+	+	+	+	+	+	+
'	'	'	'					'	'	'	'	'
'	'	'	'					'	'	'	'	'
'	'	'	'					'	'	'	'	'
m-2						0	c	+	+	+	+	+
m-1					- - - - -		0	c	+	+	+	+
m					- - - - -			0	+	+	+	+
m+1					- - - - -				0		+	+
m+2					- - - - -					0	rc	+
m+3	rc	+	+	+	- - - - -	+	+	+	+	+	+	+
m+4	+	+	+	+	- - - - -	+	+	+	+	+	+	+

= D

The subarrays of the standard form have the following forms:

[ ] denotes a subarray each of whose elements may be zero or plus

[0] denotes a square subarray all of whose main diagonal elements are zeros

[r] denotes a subarray with at least one zero in each row

[c] denotes a subarray with at least one zero in each column

[rc] denotes a subarray with at least one zero in each row and each column

[+] denotes a subarray with no zeros

In this form, the zeros are all contained within the columns passing through D_{11}, D_{22}, ---, D_{mm} and the rows passing through $D_{m+1\,m+1}$ and $D_{m+2\,m+2}$. Consequently, if unity is added to each element of each of these lines, in turn, the new matrix will have all elements positive; hence, unity can next be subtracted from every element of the matrix without producing any negative element. The new matrix can then be permuted again to standard form. Since more was subtracted than was added, to the matrix as a whole, this process must eventually terminate; it does so when the zero elements fill the main diagonal in the standard form.

The Hungarian Method proceeds by two kinds of moves. One type of move involves finding a permutation that produces a standard form, so that the lines are identified for the additions to be made. The other type of move consists in adding and subtracting to this set of lines until no further such unit changes can be made.

Following Kuhn's [7] terminology, we shall say that two zeros are "independent" if they are not in the same line. Also, a set of lines is called a "cover" if every zero is in some line of the set. A "minimal cover" is one containing the fewest lines. By the König-Egerváry theorem, the maximal number of independent zeros is equal to the number of lines in a minimal cover. For further convenience, we shall call the number of elements in a set the "cardinal" of the set.

We shall use the phrase "trial set" to denote any set of independent zeros chosen so that every zero of the matrix is in a line with at least one of the zeros of the trial set; a "trial zero" is any zero in a trial set. A "maximal trial set" is one containing the largest possible number of independent zeros. The phrase "trial cover" refers to a cover obtained using a specific trial set, and in the specific way described next.

Form the trial cover, for a given trial set, by steps as follows:

1. Include in the trial cover each row containing a trial zero, that also includes a zero (called a "row candidate") in some column containing no trial zero. Next, include in the trial cover each column containing a trial zero, that also includes a zero (called a "column candidate") in some row containing no trial zero. Consider these rows and columns in the trial cover deleted from the matrix. Repeat this step until no further lines are added to the trial cover in this way.

2. Include in the trial cover each row containing a trial zero, in the matrix remaining after the deletions of Step 1.

It follows easily, by an inspection of the standard form, that the trial cover is a minimal cover if the trial set is a maximal trial set; also, if the cardinal of the trial cover exceeds the cardinal of the trial set then the trial set is not maximal.

The Hungarian Method proceeds by steps as follows:

a. Choose a trial set. If the order of this trial set is equal to the order of the matrix then the required permutation is the one that would permute the rows so as to place these zeros on the main diagonal. If not, go next to Step b.

b. Choose a trial cover. If the cardinal of the trial cover is less than the order of the matrix then add a positive constant to all lines in the trial cover and subtract this constant from all rows of the matrix, choosing as this constant the largest integer that will leave all elements of the matrix nonnegative; we call this constant the "fuse" for the cover and go next to Step c. If the cardinal of the trial cover is not less than the order of the matrix, go next to Step d.

c. Add to the trial set any new zeros that are eligible. If the cardinal of the new trial set is less than the order of the matrix then return to Step b, otherwise the process ends as in Step a.

d. The trial set is not maximal. Delete a zero from the trial set that is on two lines of the trial cover; such zeros are called "degenerates." Add to the trial set from the candidate set, forming a new trial set, using row candidates in the row through the deleted zero and the column candidates in the column through it. Delete all trial zeros on line with any candidate added to the trial set. For each trial zero deleted in this way, add to the trial set any candidates that yielded lines (in the trial cover) containing the deleted zero. If the cardinal of the

new trial set is less than the order of the matrix then return to Step b, otherwise the process ends as in Step a.

This brief description of the Hungarian Method, including one specific method for finding trial sets and trial covers, is intended only to indicate the general nature of the method. A more detailed and precise description of the steps actually used, in the variant presented in this paper, will be given in following sections; the variant actually used differs somewhat from the one described above, and choices left arbitrary above are made entirely definite in the computer code, but the differences are not significant for present purposes. Formal mathematical proofs of the validity of the steps are not given since their validity is easily established using the "standard form" for a (0, +) matrix.

INITIAL PROBLEM PREPARATION

It would be possible to solve an assignment problem, or a transportation problem, by use of the Hungarian Method from beginning to end. However, there are a number of improvements that can be made easily; several of these are discussed in the present paper and used to advantage in the IBM 704 code. We first list some of these, and describe them briefly; then principal features will be illustrated by simple numerical examples.

Compression Subroutine

Rows of the matrix having the same pattern of zero entries are combined, as are columns having the same pattern. Except when determining fuses, compression leaves the procedures unchanged.

Reduction Subroutine

Each compressed line[3] is made to have at least as many zeros as its frequency requires. Thus, if a compressed line includes X of the original rows then it is made to have zeros in compressed columns that include at least X of the original columns.

Isolation Subroutine

The trial set is chosen by first selecting "isolated" zeros that stand alone in some compressed line. One or both of the compressed lines through such an isolate are then deleted, after their zeros are exhausted, and new isolates are added to the trial set until no isolate remains.

Shortening Subroutine

If a compressed line has no more zeros than its frequency requires, then all these zeros are added to the trial set.

Zero Subroutine

A zero is chosen in a compressed line that contains zeros in the fewest perpendicular lines. When several such zeros exist, perhaps even in more than one compressed line, that one is chosen which has the most lines parallel to it through zeros in the compressed line perpendicular to it; if a tie still remains, the zero is chosen arbitrarily from among tied cases.

These Subroutines yield a trial set and also reduce the elements of the original matrix.

---

[3] A "line" is a row or a column.

The Hungarian Method proceeds by two kinds of moves. One type of move involves finding a permutation that produces a standard form, so that the lines are identified for the additions to be made. The other type of move consists in adding and subtracting to this set of lines until no further such unit changes can be made.

Following Kuhn's [7] terminology, we shall say that two zeros are "independent" if they are not in the same line. Also, a set of lines is called a "cover" if every zero is in some line of the set. A "minimal cover" is one containing the fewest lines. By the König-Egerváry theorem, the maximal number of independent zeros is equal to the number of lines in a minimal cover. For further convenience, we shall call the number of elements in a set the "cardinal" of the set.

We shall use the phrase "trial set" to denote any set of independent zeros chosen so that every zero of the matrix is in a line with at least one of the zeros of the trial set; a "trial zero" is any zero in a trial set. A "maximal trial set" is one containing the largest possible number of independent zeros. The phrase "trial cover" refers to a cover obtained using a specific trial set, and in the specific way described next.

Form the trial cover, for a given trial set, by steps as follows:

1. Include in the trial cover each row containing a trial zero, that also includes a zero (called a "row candidate") in some column containing no trial zero. Next, include in the trial cover each column containing a trial zero, that also includes a zero (called a "column candidate") in some row containing no trial zero. Consider these rows and columns in the trial cover deleted from the matrix. Repeat this step until no further lines are added to the trial cover in this way.

2. Include in the trial cover each row containing a trial zero, in the matrix remaining after the deletions of Step 1.

It follows easily, by an inspection of the standard form, that the trial cover is a minimal cover if the trial set is a maximal trial set; also, if the cardinal of the trial cover exceeds the cardinal of the trial set then the trial set is not maximal.

The Hungarian Method proceeds by steps as follows:

a. Choose a trial set. If the order of this trial set is equal to the order of the matrix then the required permutation is the one that would permute the rows so as to place these zeros on the main diagonal. If not, go next to Step b.

b. Choose a trial cover. If the cardinal of the trial cover is less than the order of the matrix then add a positive constant to all lines in the trial cover and subtract this constant from all rows of the matrix, choosing as this constant the largest integer that will leave all elements of the matrix nonnegative; we call this constant the "fuse" for the cover and go next to Step c. If the cardinal of the trial cover is not less than the order of the matrix, go next to Step d.

c. Add to the trial set any new zeros that are eligible. If the cardinal of the new trial set is less than the order of the matrix then return to Step b, otherwise the process ends as in Step a.

d. The trial set is not maximal. Delete a zero from the trial set that is on two lines of the trial cover; such zeros are called "degenerates." Add to the trial set from the candidate set, forming a new trial set, using row candidates in the row through the deleted zero and the column candidates in the column through it. Delete all trial zeros on line with any candidate added to the trial set. For each trial zero deleted in this way, add to the trial set any candidates that yielded lines (in the trial cover) containing the deleted zero. If the cardinal of the

new trial set is less than the order of the matrix then return to Step b, otherwise the process ends as in Step a.

This brief description of the Hungarian Method, including one specific method for finding trial sets and trial covers, is intended only to indicate the general nature of the method. A more detailed and precise description of the steps actually used, in the variant presented in this paper, will be given in following sections; the variant actually used differs somewhat from the one described above, and choices left arbitrary above are made entirely definite in the computer code, but the differences are not significant for present purposes. Formal mathematical proofs of the validity of the steps are not given since their validity is easily established using the "standard form" for a (0, +) matrix.

INITIAL PROBLEM PREPARATION

It would be possible to solve an assignment problem, or a transportation problem, by use of the Hungarian Method from beginning to end. However, there are a number of improvements that can be made easily; several of these are discussed in the present paper and used to advantage in the IBM 704 code. We first list some of these, and describe them briefly; then principal features will be illustrated by simple numerical examples.

Compression Subroutine

Rows of the matrix having the same pattern of zero entries are combined, as are columns having the same pattern. Except when determining fuses, compression leaves the procedures unchanged.

Reduction Subroutine

Each compressed line[3] is made to have at least as many zeros as its frequency requires. Thus, if a compressed line includes X of the original rows then it is made to have zeros in compressed columns that include at least X of the original columns.

Isolation Subroutine

The trial set is chosen by first selecting "isolated" zeros that stand alone in some compressed line. One or both of the compressed lines through such an isolate are then deleted, after their zeros are exhausted, and new isolates are added to the trial set until no isolate remains.

Shortening Subroutine

If a compressed line has no more zeros than its frequency requires, then all these zeros are added to the trial set.

Zero Subroutine

A zero is chosen in a compressed line that contains zeros in the fewest perpendicular lines. When several such zeros exist, perhaps even in more than one compressed line, that one is chosen which has the most lines parallel to it through zeros in the compressed line perpendicular to it; if a tie still remains, the zero is chosen arbitrarily from among tied cases.

These Subroutines yield a trial set and also reduce the elements of the original matrix.

---

[3] A "line" is a row or a column.

Compression Example —

	3	2	1	4	5
2	2	0	3	1	5
3	4	0	1	0	2
3	1	3	2	0	4
3	3	0	0	2	1
4	5	0	1	2	3

→

	8	2	1	4	
6	+	0	+	+	1, 5
3	+	0	+	0	
3	+	+	+	0	
3	+	0	0	+	
	1, 5				

The first and fifth rows, with Frequencies 2 and 4, compress into a new first row with Frequency 6; similarly the first and fifth columns compress.

Reduction Example — The first compressed row requires 6 zeros but has only 2; similarly the first compressed column requires 8 zeros and has none. The smallest amount that can be subtracted from the first compressed row, and yield a new zero, is 1 (the fuse); to retain the old zero in the first compressed row and second column, we must add 1 to the second column.

	3	2	1	4	5	
2	2	0	3	1	5	-1
3	4	0	1	0	2	
3	1	3	2	0	4	
3	3	0	0	2	1	
4	5	0	1	2	3	-1
		+1				

=

1	0	2	0	4	
4	1	1	0	2	
1	4	2	0	4	
3	1	0	2	1	
4	0	0	1	2	

→

	8	2	1	4	
2	+	0	+	0	
6	+	+	+	0	2,3
3	+	+	0	+	
3	+	0	0	+	
	1,5				

Compressed Rows 2 and 3, and compressed Column 1, all require reduction; consider compressed Column 1 next, with Fuse 1.

	3	2	1	4	5
2	1	0	2	0	4
3	4	1	1	0	2
3	1	4	2	0	4
3	3	1	0	2	1
4	4	0	0	1	2
	-1				-1

=

	3	2	1	4	5
2	0	0	2	0	3
3	3	1	1	0	1
3	0	4	2	0	3
3	2	1	0	2	0
4	3	0	0	1	1

Here, no compression is possible and only Row 5 and Column 5 require reduction; consider Row 5 next, with Fuse 1.

	3	2	1	4	5															
2	0	0	2	0	3			0	1	3	0	3		5	0	+	0	+	1,3
3	3	1	1	0	1			3	2	2	0	1		3	+	+	0	+	
3	0	4	2	0	3	=		0	5	3	0	3	→	3	+	+	+	0	
3	2	1	0	2	0			2	2	1	2	0		4	+	0	0	0	
4	3	0	0	1	1	-1		2	0	0	0	0				2,3			
		+1	+1																

No further compression or reduction is possible, so the reduction phase is completed.

Isolation Subroutine — The trial set is obtained by choosing isolated zeros (in the compressed matrix) as shown next, where the entry replacing the zero is the frequency for that trial zero and the subscripts indicate the order in which the isolated zeros were chosen.

	0	0	4	2			0	0	0	1
2	3_2	+	0	+		0	3_2	+	2_4	+
3	+	+	0	+	=	1	+	+	2_5	+
0	+	+	+	3_1		0	+	+	+	3_1
1	+	3_3	0	0		0	+	3_3	0	1_6

Note that lines are treated as though they were deleted after their frequency goes to 0.

This example does not illustrate the use of recompression during our choice of the trial set, but recompression normally occurs several times during the process as lines are deleted.

Shortening Subroutine — When the reduced and compressed matrix reaches a stage where no further isolated zeros are found, perhaps after several deletions of lines during construction of a trial set, then the Shortening Subroutine may add zeros to the trial set, as indicated in the following example.

	5	2	3			0	0	0
5	+	0	0		0	+	2	3
2	0	0	+	=	0	2	0	+
3	0	+	0		0	3	+	0

Zero Subroutine — When reduced and compressed matrix reaches a stage where further application of the Isolation and Shortening Subroutines yield no further trial zeros, then the Zero Subroutine takes effect. The following example illustrates this process.

	2	3	4			0	3	4
5	+	0	0		5	+	0	0
2	0	0	+	=	2	0	0	+
2	0	+	0		0	2	+	0

The trial zero in Row 3 and Column 1 is chosen because it is in the line having fewest zeros, namely 4 in Column 1, and is preferred over the zero in Row 2 because Row 3 has more zeros, with 6, than does Row 2, with only 5.

Fuse Rule

The Reduction Subroutine is usually improved if each uncompressed line is first reduced enough in one step to yield the required number of zeros; this will sometimes mean the use of more than one fuse in a single step. The fuses are selected according to the "Fuse Rule," which states simply that just enough fuses are selected, in increasing order of numerical size, to yield the required number of zeros for the line being reduced. The following example illustrates application of the Fuse Rule.

	3	2	1	4	5				3	2	1	4	5
2	1	0	2	0	4			2	0	0	2	0	2
3	4	1	1	0	2			3	3	1	1	0	0
3	1	4	2	0	4		=	3	0	4	2	0	2
3	3	1	0	2	1	+1		3	3	2	1	3	0
4	4	0	0	1	2			4	3	0	0	1	0
	-1				-2								

The fuse for Column 1 is 1, because Rows 1 and 3 yield 5 zeros to meet the requirement of 3. The fuse for Column 5 is 2, because Row 4 (and fuse = 1) does not yield the 5 zeros required for Column 5; the next larger entry in Column 5 is 2, and this becomes the actual fuse used when the Fuse Rule is applied to Column 5. Actually, for this example, our reduction using the Fuse Rule has completed the problem, except for finding the trial set which then proves to be a solution.

Many problems are completely solved by the first trial set, after the initial preparation. When they are not completely solved, the reduction in the matrix is large enough to make a substantial cut in computational time required to complete the solution by some other procedure. For example, the Simplex Method could be used after this initial preparation with the trial set a part of the first trial basis. We shall return next to the Hungarian Method, and illustrate the use of the trial cover by a numerical example.

TRIAL COVER EXAMPLE

We use the Isolation Subroutine example to illustrate the procedure for forming a trial cover. The trial set is shown as numerical entries in the (compressed) matrix, where a 0 denotes a zero in the (compressed) matrix that is a zero not in the trial set.

	3	3	4	5
5	3	+	2	+
3	+	+	2	+
3	+	+	+	3
4	+	3	0	1

We first note that the zeros in positions (3, 4) and (4, 4) yield row candidates, and this leads immediately to the inclusion of (compressed) Rows 3 and 4 in the trial cover; this is so because (compressed) Column 4 includes five lines but only four trial zeros. Similarly, the zero in position (2, 3) yields a column candidate, and places Column 3 in the trial cover. After deleting (compressed) Rows 3 and 4, and (compressed) Column 3, the following matrix remains.

	3	3	5
5	3	+	+
3	+	+	+

In this matrix, the zero in position (1, 1) yields a column candidate, and the first column is added to the trial cover; thus, compressed Column 1 of the original matrix is in the trial cover. No zeros remain, after deleting Column 1, so the trial cover is complete; it consists of compressed lines (R3, R4, C1, C3). Note that the cardinal of the trial cover is 14, and is equal to the cardinal of the trial set, so it follows that the trial set is maximal.

Our next example shows how the trial cover is formed when the trial set is not maximal.

	3	3	4	5
5	3_c	+	0_c	+
3	+	+	0_c	+
3	+	+	+	3_r
4	+	0_r	4	0_r

The trial cover includes compressed lines (R3, R4, C1, C3), with cardinal 14, while the cardinal of the trial set is 10. The candidate zeros are marked with an r or c as subscript, to indicate whether row or column candidate. The zero in position (4, 3) may be deleted from the trial set because it lies in intersecting paired lines of the trial cover; after its removal the trial set is increased by using candidate zeros in lines R4 and C3.

FORTRAN CODES

Two separate FORTRAN II codes were written and tested, based on the general principles of the algorithm. The Algorithm Flow Diagram (Figure 1) shows the steps in the computation. The earlier of these, called "CODE I," used recompression at intermediate stages of Step 1B, but the latter one, called "CODE II," did not. There were other important differences between CODES I and II, and neither code proceeds exactly as in the flow diagram for the algorithm. We shall discuss only CODE II in the present paper; results with CODE I were reported previously [11].

CODE II was written to include provision for measuring the computing time required, in any particular run, for each of several stages within the whole run. This was made possible by the use of an internal clock that could be read and recorded, at any point during the calculation, when required by the code. Provision was made for reading the clock at 12 such points in the code, including 7 clock readings at points normally passed more than once during a problem; these repeated clock readings can be made the first 20 times the point is reached, or

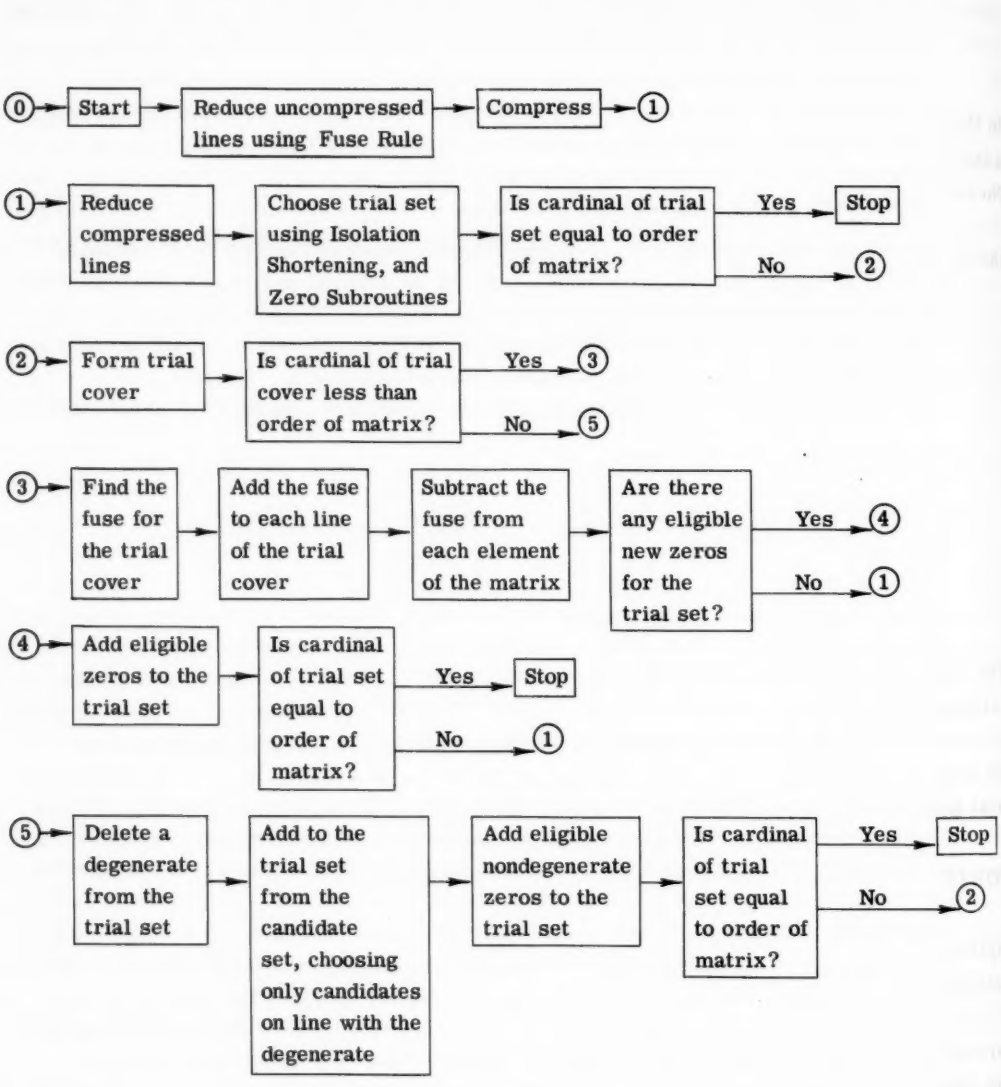

Figure 1 - Algorithm flow diagram

fewer at each point if desired. Clock readings may also be taken at the start and end of the computing cycle on a problem, and are always taken at the beginning of input and end of output for a problem.

CODE II also provides for a count of the number of times each of 17 points is reached during the calculation of each problem.

CODE II includes a subroutine that may be called in to compose a problem of any desired size, up to the dimensional limit of 60 x 150 for the cost matrix. The cost matrix, and the row and column frequencies, are generated by the subroutine using a pseudorandom number technique.

Although CODE II was written and compiled to handle a transportation problem having at most 60 rows and 150 columns, it would be quite a simple matter to recompile the code to handle any problem having fewer than 24,000 elements in the cost matrix.

Other special features of CODE II, that will not be discussed in any detail in this paper, include provision for selecting among a few alternative algorithms and several options on control of output data.

Figure 2 presents a condensed flow diagram for the CODE II FORTRAN Subroutines, and shows where CLOCK and INDICATOR readings are taken during a computation. The FORTRAN Flow Diagram (Figure 2) shows the computational steps in some detail. This will be useful for our subsequent discussion of times and frequencies through various computing stages of actual problems run.

The FORTRAN method for introducing CLOCK and INDICATOR readings when desired during the run is illustrated by the following example; this example is taken from the start of SUBROUTINE P13.

```
      SUBROUTINE P13
1     IND(15) = IND(15) + 1
2     IF(KEY(31)) 7, 7, 3
3     IF(IND(15)-KEY(19)) 4, 4, 7
4     CALL CLOCK (IND(51))
5     L = IND(15)
6     LOCK(L,5) = IND(51) - LOCK(1,1)
7
```

Instruction 1 increases INDICATOR 15 by unity each time SUBROUTINE P13 is called; since IND(15) is initialized to zero, at the start of the run, this yields the frequency of use of SUBROUTINE P13.

Instruction 2 constitutes an option in reading the internal clock at this stage of the run. KEY(31) is data read in at the start of the run. If KEY(31) > 0 then the clock is read, otherwise not.

Instruction 3 constitutes an option on the number of times the internal clock is read at this stage. KEY(19) is data read in at the start of the run, and can be any integer less than 21. If KEY(19) ≤ 0 the clock is not read, if $0 <$ KEY(19) ≤ 20 the clock is read the first KEY(19) times it reaches this stage, and KEY(19) > 20 is not allowed.

Instruction 4 calls SUBROUTINE CLOCK(K), which was a library subroutine at the IBM Computation Center in Poughkeepsie that would stop the computation while an internal clock was stopped and read to the nearest 0.01 minutes. The clock reading was then stored in location K. In CODE II the location IND(51) was used for K. The result of this instruction is

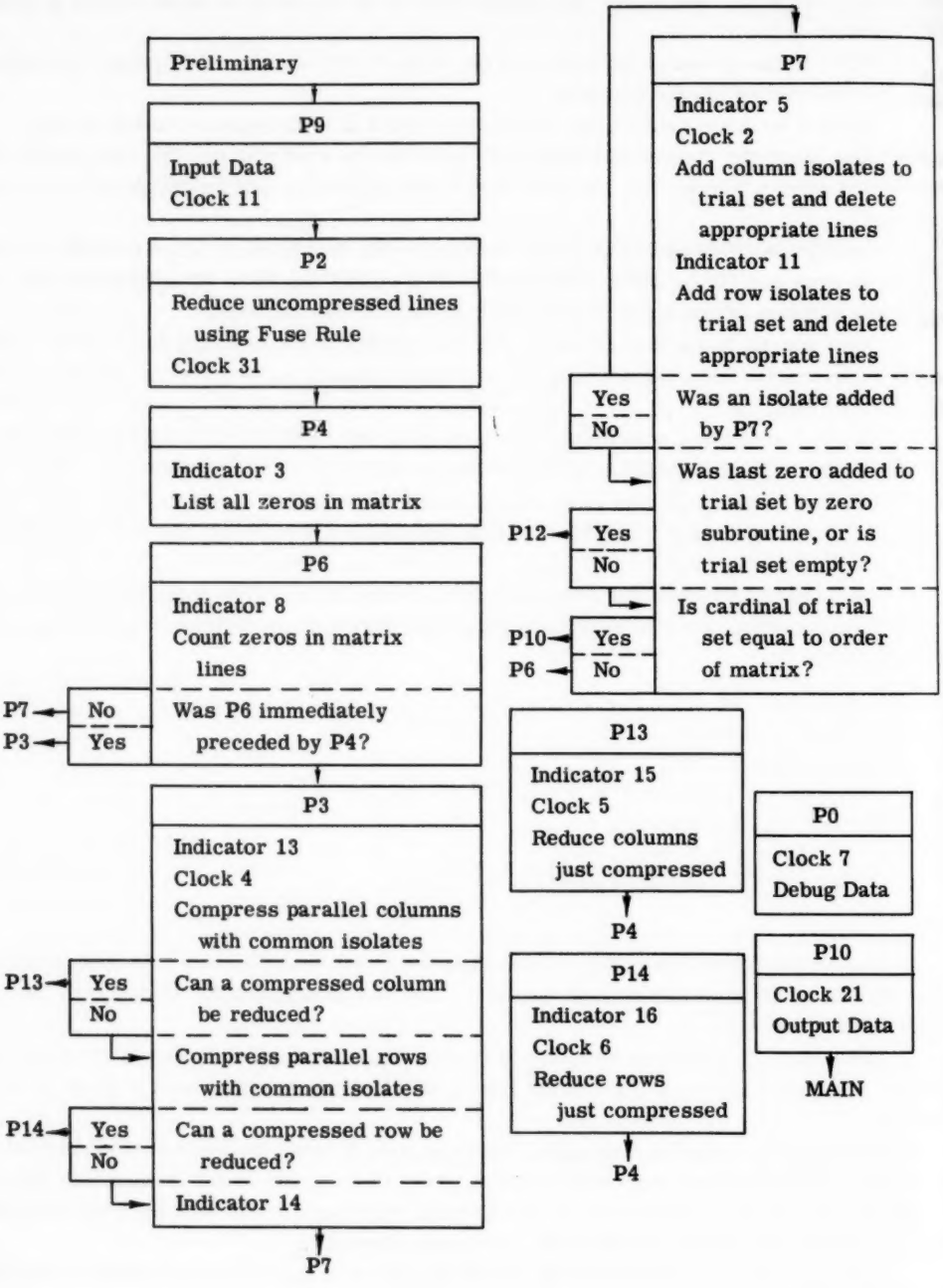

Figure 2 - FORTRAN flow diagram

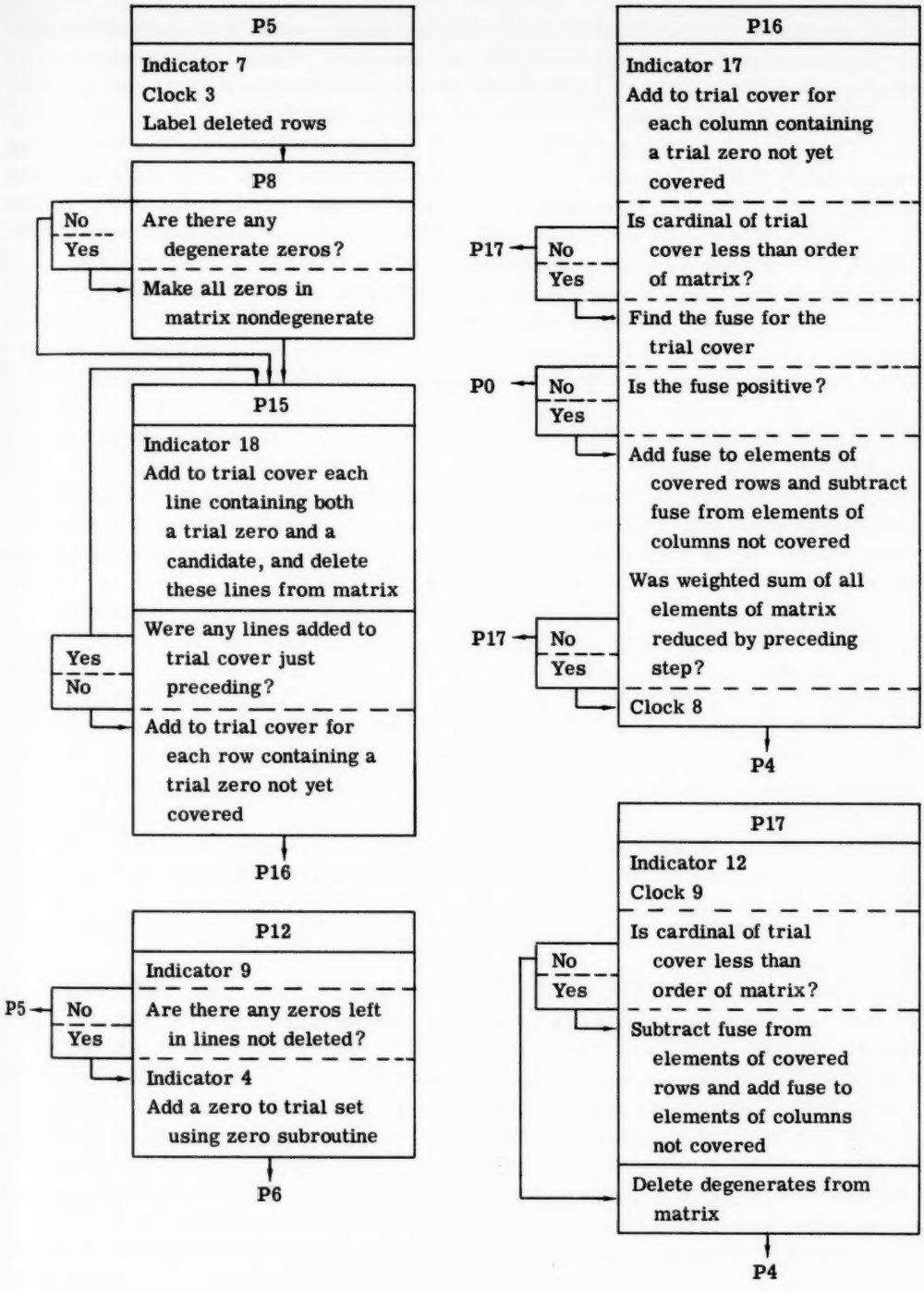
P5
Indicator 7
Clock 3
Label deleted rows
P8
No
Yes
Are there any degenerate zeros?
Make all zeros in matrix nondegenerate
P15
Indicator 18
Add to trial cover each line containing both a trial zero and a candidate, and delete these lines from matrix
Yes
No
Were any lines added to trial cover just preceding?
Add to trial cover for each row containing a trial zero not yet covered
P16
P12
Indicator 9
P5
No
Yes
Are there any zeros left in lines not deleted?
Indicator 4
Add a zero to trial set using zero subroutine
P6
P16
Indicator 17
Add to trial cover for each column containing a trial zero not yet covered
P17
No
Yes
Is cardinal of trial cover less than order of matrix?
Find the fuse for the trial cover
P0
No
Yes
Is the fuse positive?
Add fuse to elements of covered rows and subtract fuse from elements of columns not covered
Was weighted sum of all elements of matrix reduced by preceding step?
P17
No
Yes
Clock 8
P4
P17
Indicator 12
Clock 9
No
Yes
Is cardinal of trial cover less than order of matrix?
Subtract fuse from elements of covered rows and add fuse to elements of columns not covered
Delete degenerates from matrix
P4

Figure 2 (Cont'd) - FORTRAN flow diagram

then to change the value at IND(51) to the current clock time, and while the clock was stopped so that clock reading time was excluded from computing time. Since it takes about 0.01 minutes to take one clock reading the SUBROUTINE CLOCK was used sparingly.

Instruction 5 is a coding technicality that replaces IND(15) by L.

Instruction 6 calculates the elapsed computing time, in hundredths of minutes, from the start of SUBROUTINE P2 until the present; it also stores this value at LOCK (L,5). In general, LOCK(L,I) will be the time since starting P2 to reach the stage I the Lth time; in P13 we have I = 5. Since LOCK(1,1) was used for the time SUBROUTINE P2 was first called, and since IND(51) is the present time, this instruction sets LOCK(L,5) equal to the elapsed time.

Instruction 7 starts the sequence of calculations done in SUBROUTINE P13.

The output data include the matrix LOCK(L,I), for L = 1, 2, ---, KEY(19) $\leq$ 20 and I = 1, 2, ---, 20. The vector IND(J), for J = 1, 2, ---, 60 is also shown in the output data. Clock times and subprogram frequencies, of this kind, constitute the data presented in this paper.

FORTRAN FLOW DIAGRAM

The name of the FORTRAN II Subroutine appears first in each box of the FORTRAN flow diagram (Figure 2). CLOCK X entries show the points during the computation at which clock readings may be taken. INDICATOR Y entries show the points during the computation at which counts are made of the frequencies of passing these points.

CODE II is represented correctly by this FORTRAN Flow Diagram, but differs in several respects from the Algorithm Flow Diagram. For example, the matrix is not recompressed at each step in CODE II that is indicated by the Algorithm Flow Diagram, and several other important changes have been made in CODE II to facilitate the use of FORTRAN.

There is also a coding defect in CODE II, but this defect results only in stopping a few problems before the computation is entirely completed. CODE II is arranged so that this case results in a special output, through Subroutine P0, that provides the data necessary to complete the problem easily. This output, like several others included in CODE II, automatically indicates the point of error whether due to machine difficulty or some other unexpected trouble. These possible sources of error will be ignored throughout the rest of our discussion, since they are rare and recognizable.

COMPUTATIONAL RESULTS

We present our computation results first in terms of CLOCK times. For example, T = CLOCK 21-CLOCK 11 represents actual total computing time; if SUBROUTINE P0 is used, then actual total computing time is T = CLOCK 7-CLOCK 11. As another example, T_{83} = CLOCK 8-CLOCK 3 represents actual computing time used in going once through SUBROUTINES P5, P8, P15, P16 if CLOCK 9 was not read meanwhile. This latter example illustrates one way in which individual CLOCK readings may be used to trace cycle times in various portions of the complete run. Similarly, if CLOCK 4 follows CLOCK 9 (or CLOCK 5, 6, 8 or 31) then SUBROUTINE P3 (rather than SUBROUTINE P7) followed SUBROUTINE P6; in this general manner the computing time taken for various segments of the run can be determined.

CODE II is written so that intermediate outputs can be obtained during a run in very many different ways, but this feature will not be discussed in the present paper. Suffice it to say that this feature permits the use of INDICATOR readings, in conjunction with CLOCK readings, to follow the path and timing of computational steps in considerable detail when appropriate intermediate outputs are available.

We turn now to CLOCK results on one particular problem (#E116.1), in order to show the kind of timing data obtained. This is the only problem that has been computed also by a competing code, called IB-TFL, perhaps the fastest standard IBM 704 code (SHARE 464)[4] available for the transportation problem. This problem was a 29 x 116 pseudorandom transportation problem that took 3.03 minutes with CODE II, in comparison with 3.17 minutes with IB-TFL; this difference is so small that the two codes may be considered equally efficient for this one 29 x 116 problem. On a second pseudorandom problem (#E116.2), of size 29 x 116, CODE II took 1.90 minutes.

The beginning of the flow through #E116.1 may be followed by noting CLOCK readings, as presented in Table 1. In other words, the initial uncompressed reduction through P2 took 0.24 minutes. Then the P4 → P6 → P3 → P13 cycle was repeated 9 times in 0.34 minutes, to complete the compressed isolate columns reduction. Next, the P7 cycle was repeated 6 times, in 0.02 minutes, to list the isolated zeros. Next the P12 → P5 → P8 → P15 → P16 sequence took 0.05 minutes to find a trial cover, fuse, and reduction before returning to P4 → P6 → P7. Since we had arranged to have only 9 readings on CLOCK 2, we cannot trace the path through all the subroutines after 0.72 minutes. We do obtain the first 10 times at which CLOCKS 3 and 8 are read and, from these data, it is seen that: the P5 → P8 → P15 → P16 portion takes about 0.06 minutes each time, and the portion from P16 back to P5 through P4 → P6 → P7 → P12 takes about 0.09 minutes each time.

TABLE 1

CLOCK Readings for #E116.1

Time	0	24	25	25	26	26	28	28	30	31	33	34	36	37	39	40	etc.
Clock	11	31	4	5	4	5	4	5	4	5	4	5	4	5	4	5	
Subroutine	P9	P2	P3	P13	P3	P13	P3	P13	P3	P13	P3	P13	P3	P13	P3	P13	

Time	58	60	60	60	60	60	60	65	70	72	72	etc.	73	79	88	95	etc.
Clock	2	2	2	2	2	2	3	8	2	2	2		3	8	3	8	
Subroutine	P7	P7	P7	P7	P7	P7	P5	P16	P7	P7	P7		P5	P16	P5	P16	

The indicators not only show frequencies with which each of several points are passed during the computation, but also may be used to assist in tracing the actual path of the computation through the run. The Indicator readings for problem #E116.1 are presented in Table 2. For example, since I12 = I16 = 0 it is seen that P14 and P17 were not used; also since I4 = 0, the segment P12 → P6 was never used. By continuing in this way, we find that the paths used were as presented in Table 3. This shows that Subroutine P7 was entered 99 times; of these, 15 led to P12, 15 to P6, 68 back through P7, and 1 to Stop. It shows also that Subroutine P7 was entered 15 times from P6, 16 from P3, and 68 from itself.

The 24-path frequencies can always be determined by findings entries for the 24 cells marked x in Figure 3 that yield row and column sums equal to the observed Indicator values.

[4]SHARE 464 is a code based upon the Ford-Fulkerson [12] network flow algorithm, which is another variant of Kuhn's Hungarian Method.

617404 O - 61 - 5

TABLE 2

Indicator Readings for #E116.1

Indicator #	3	8	13	14	15	16	5	9	4	7	18	17	12
Frequency	27	42	27	16	11	0	99	15	0	15	79	15	0
Subroutine	P4	P6	P3		P13	P14	P7	P12		P5	P15	P16	P17

TABLE 3

Path Frequencies for #E116.1

Indicator:	3	5	7	8	9	13	14	15	17	18	Stop	Total
Start	1											1
3				27								27
5		68		15	15						1	99
7										15		15
8		15				27						42
9			15									15
13							16	11				27
14		16										16
15	11											11
17	15											15
18									15	64		79
Total	27	99	15	42	15	27	16	11	15	79	1	347

Other pseudorandom problems were generated and solved by CODE II, with results as presented in Table 4. All the pseudorandom problems were generated so that the elements of the cost matrix were non-negative integers less than 10,000. The rim totals were positive integers smaller than 10,000, and adjusted so that the row sum and column sum were equal and smaller than 30,000. Thus, they might be called transportation problems.

Several other small problems were run, for various tests, each of which was constructed for some particular purpose. Among these were one 9x9 assignment problem (unit rim totals) and 30 5x15 assignment problems (unit column rim totals). The 30 5x15 assignment problems each took less than 0.02 minutes computing time, except for one that was not quite completed—due to the same defect in CODE II that kept the 58x145 problems from finishing. The 9x9 assignment problem was constructed carefully to stretch CODE II, and did in fact require 0.26 minutes computing plus system time.

Indicator #	3	4	5	7	8	9	12	13	14	15	16	17	18	Stop
Start	x													
3					x									
4					x									
5			x		x	x								x
7													x	
8			x					x						
9		x		x										
12	x													
13									x	x	x			
14			x											
15	x													
16	x													
17	x						x							x
18												x	x	

Figure 3 - Possible paths

TABLE 4

Computing Times

Dimension	Number Solved	Median Computing Time	Smallest Computing Time	Largest Computing Time	Input Output Time
		← In minutes →			
58 x 145	2	13.56*	12.72*	14.41*	?
29 x 116	2	2.47	1.90	3.03	.18
29 x 29	3	.55	.49	.85	.14
20 x 29	13	.25	.18	.37	.14
9 x 10	7	.45†	.13†	.65†	?

*Actually, these problems were not quite completed, but the output is adequate for easy completion.
†All times too high because intermediate outputs were obtained.

The Indicator values and computing times for several problems are presented in Table 5. Problems E29.1 through E29.12 were run using CODE II controlled so as to omit Subroutines P3, P13 and P14, contrary to the Flow Diagram of Figure 2. The 24-path frequencies can be computed easily, for any of these runs, by filling in the 24 marked cells of Figure 3.

TABLE 5

Indicator Readings and Computing Times

Problem		Time Min	Indicator Number													Total
Size	Number		3	4	5	7	8	9	12	13	14	15	16	17	18	
29 x 116	E116.1	3.03	27	0	99	15	42	15	0	27	16	11	0	15	79	346
"	E116.2	1.90	16	10	99	8	44	18	0	16	9	7	0	8	46	281
58 x 145	E145.1	14.41	48	33	432	27	141	60	0	48	27	21	0	26	239	1102
"	E145.2	12.72	43	15	302	23	96	38	0	43	23	19	1	23	193	819
20 x 29	E29.1	0.37	10	1	58	9	21	10	0	0	0	0	0	9	38	156
"	E29.2	0.23	11	0	56	10	21	10	0	0	0	0	0	10	38	156
"	E29.3	0.25	11	0	59	10	21	10	0	0	0	0	0	10	42	163
"	E29.4	0.30	13	0	64	12	25	12	0	0	0	0	0	12	46	184
"	E29.5	0.29	14	0	75	13	27	13	0	0	0	0	0	13	49	204
"	E29.6	0.26	12	0	61	11	23	11	0	0	0	0	0	11	41	170
"	E29.7	0.24	11	0	57	10	21	10	0	0	0	0	0	10	42	161
"	E29.8	0.24	10	0	46	9	19	9	0	0	0	0	0	9	36	138
"	E29.9	0.18	8	0	40	7	15	7	0	0	0	0	0	7	27	111
"	E29.10	0.26	9	9	89	8	35	17	0	0	0	0	0	8	43	218
"	E29.11	0.28	12	0	59	11	23	11	0	0	0	0	0	11	48	175
"	E29.12	0.25	11	0	56	10	21	10	0	0	0	0	0	10	38	156
"	E29.12	0.25	11	0	56	10	21	10	0	1	1	0	0	10	38	158
"	E29.13	0.36	16	6	115	14	42	20	0	1	0	1	0	14	65	294
29 x 29	ES29.1	0.85	26	0	151	21	47	21	0	26	22	4	0	21	110	449
"	ES29.2	0.55	22	0	89	15	37	15	0	22	16	6	0	15	63	300
"	ES29.3	0.49	16	0	85	14	30	14	0	16	15	1	0	14	63	268

Test runs with the recompression subroutines (P3, P13, P14) omitted in CODE II have shown that their inclusion shortens computation times appreciably for large problems. Experience with CODE I, where complete recompression was done at each step rather than recompression only of lines having common isolates as in CODE II, indicates the desirability of more extensive recompression than is used in CODE II; however, the FORTRAN techniques used in

CODE I are not effective in accomplishing more extensive recompression. Comparative tests of other alternatives have suggested various improvements that can easily be made in CODE II, but without deviating from the essential steps of our mathematical algorithm, and data on segmental computing times have helped in choosing among some of these alternatives.

CONCLUSIONS

CODE II makes effective use of the Hungarian Method for solving transportation and assignment problems on the IBM 704. Computing experience suggests that CODE II is now as fast as the IB-TFL code, or any other existing code, and CODE II would surely be improved further if rewritten to gain efficiency and speed.

The mathematical algorithm supporting CODE II makes essential use of several reduction steps, preliminary to the execution of zero covering and selection subroutines based on the present author's proof of the König-Egerváry Theorem [10]. These preliminary reductions should be useful with other codes for the Hitchcock distribution problem, whether based upon the Hungarian Method, the Simplex Method, or any other method.

The techniques employed in CODE II, for observing timing and frequencies of passages through segments of a particular computation, are very useful in determining choices among alternative codes. Experience with CODE II has provided timing and frequency data that shows where further substantial improvements may possibly be made in CODE II, while holding to the same mathematical algorithm.

REFERENCES

[1] Merrill M. Flood, "The Traveling Salesman Problem," Operations Research, Vol. 4, No. 1, February 1956, pp. 61-75.

[2] C. W. Churchman, R. L. Ackoff, and E. L. Arnoff, "Introduction to Operations Research," Wiley, Chapters 11 and 12 (1957).

[3] George B. Dantzig, "Application of the Simplex Method to a Transportation Problem," In: T. C. Koopmans (Ed.), Activity Analysis of Production and Allocation, Wiley (1951).

[4] Frank L. Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities," Journal of Mathematics and Physics, Vol. 20 (1941), pp. 224-230.

[5] L. Kantorovich, "On the Translocation of Masses," Management Science, Vol. 5, No. 1, October 1958, pp. 1-4.

[6] Merrill M. Flood, "On the Hitchcock Distribution Problem," Pacific Journal of Mathematics, Vol. 3, No. 2, June 1953, pp. 369-386.

[7] H. W. Kuhn, "The Hungarian Method for the Assignment Problem," Naval Research Logistics Quarterly, Vol. 2, Nos. 1, 2, March-June 1955, pp. 83-98.

[8] Denes König, "Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre," Mathematisches Annalen, Vol. 77 (1916), pp. 443-465.

[9] J. Egerváry, "Matrixok Kombinatorius Tulajdonsagairol," Mathematikai és Fizikai Lapok, Vol. 38 (1931), pp. 16-28.

[10] Merrill M. Flood, "An Alternative Proof of a Theorem of König as an Algorithm for the Hitchcock Distribution Problem," In: R. Bellman and M. Hall, Jr. (Eds.), Proceedings of Symposia in Applied Mathematics, Vol. 10, American Mathematical Society, Providence, R. I. (1960), pp. 299-307.

[11] Merrill M. Flood, "A Transportation Code," In: Philip Wolfe (Ed.), The RAND Symposium on Mathematical Programming, The RAND Corporation (1960), pp. 79-80 (abstract).

[12] L. R. Ford, Jr., and D. R. Fulkerson, "Solving the Transportation Problem," Management Science, Vol. 3, No. 1, October 1956, pp. 24-32.

* * *

APPROXIMATELY OPTIMAL ONE-DIMENSIONAL SEARCH POLICIES IN WHICH SEARCH COSTS VARY THROUGH TIME*

Brian Gluss

Armour Research Foundation of Illinois Institute of Technology
10 West 35th Street, Chicago, Illinois

Consider a model in which there are N neighboring cells in one of which there is an object that it is required to find. The a priori probabilities of the object being in cells 1, ... , N are $p_1, \ldots, p_N$ respectively, and the costs of examination of these cells are $t_1, \ldots, t_N$ respectively; the search policy is considered to be optimal when the statistical expectation of the total cost of the search is minimized. For the case in which the t_i's are constant throughout the search, an optimal policy solution has previously been found by Bellman and by Smith. In the present paper it is assumed that the costs comprise a travel cost dependent upon the distance from the last cell examined, in addition to a fixed examination cost: initially, assuming that the searcher is next to cell 1, $t_i = i + t$, where t is constant; and from then onwards, assuming that the j^{th} cell has just been examined, $t_i = |i - j| + t$. An optimal search strategy is found in the case where the p_i's are all equal, and approximately optimal strategies in the case where p_i is proportional to i. The latter case has application to defense situations where complete searches occur at successive intervals of time, and hence the enemy objects are thinned out the nearer they come to the defense base.

INTRODUCTION

A model that has been considered by, for example, Bellman [1] and Smith [2] is one in which it is required to determine a policy that will minimize the statistical expectation of the cost of finding an object that is in one of N cells. The a priori probabilities of its being in cells 1, ... , N are $p_1, \ldots, p_N$ respectively, and the costs associated with the examination of these cells are $t_1, \ldots, t_N$ respectively. In the case in which the t_i's are constant throughout the search, an exact policy solution has been found (see [1], [2]); this comprises examining firstly the cell r for which t_r/p_r is a minimum, then replacing the p's by probabilities conditional upon the outcome of the examination, re-computing the t_i/p_i, and continuing. This of course implies that the t_r/p_r need only be computed and ordered once, since the conditional probabilities associated with the remaining unexamined cells are proportional to the a priori probabilities.

However, suppose that in defining the t_i's the cells are given a geographical relationship and that the cost of examining one cell after having examined another is made to be

*Manuscript received October 30, 1959.

partially dependent upon the distance between the two cells. Then the t_i's vary through the search, and are functions of which cell was last examined. In this article, it is assumed that the cells are in a straight line, and that t_i consists of two parts: (i) the distance $|i - j|$ between cell i and the previously examined cell j, and (ii) a fixed examination cost t. That is

$$t_i = |i - j| + t .$$

It is also assumed that at the beginning of the search the searcher is at a base next to cell 1, and that consequently $t_i = i + t$ initially.

Three models are considered: in Model I, the p_i's are assumed equal (and hence equal to 1/N), and an exact policy solution is found. In Model II, the p_i's are assumed monotonically decreasing with i, and once again an exact solution is found. In Model III, the p_i's are assumed proportional to the distance i from base (and hence equal to $2i/N(N+1)$), and an approximately optimal policy solution is found. Model III appears to be of considerable practical interest as a model applicable to submarine search strategies: the base may be equated with shore, or the place to be defended, and the p_i's seem realistic in that, if such searches were carried out periodically and sufficiently often, then the enemy would be thinned out the closer they get to shore, when they have been in the search area for a longer period of time and hence have been subjected to a larger number of searches.

MODEL I — $p_i = 1/N$, all i. Consider the order of search $(i_1, \ldots, i_N)$, where $i_1, \ldots, i_N$ is some arrangement of $1, \ldots, N$. Then the expected cost of the search is

$$\begin{aligned} E(\text{cost}) &= t + i_1 + \frac{N-1}{N}\left[t + |i_1 - i_2|\right] + \frac{N-2}{N}\left[t + |i_2 - i_3|\right] + \ldots + \frac{1}{N}\left[t + |i_{N-1} - i_N|\right] \\ &= \frac{(N+1)\,t}{2} + i_1 + \frac{N-1}{N}|i_1 - i_2| + \frac{N-2}{N}|i_2 - i_3| + \ldots + \frac{1}{N}|i_{N-1} - i_N| . \end{aligned} \tag{1}$$

Except for the first term, which is independent of the i's, each term is minimized when $i_1 = 1, |i_1 - i_2| = 1, \ldots, |i_{N-1} - i_N| = 1$. Since these minima may be attained in the arrangement $(1, 2, \ldots, N)$, it therefore follows that this is the best policy.

MODEL II — p_i monotonically decreasing with i. In this case the expected cost is

$$\begin{aligned} E(\text{cost}) &= t + i_1 + \left(1 - p_{i_1}\right)\left[t + |i_1 - i_2|\right] + \left(1 - p_{i_1} - p_{i_2}\right)\left[|t + i_2 - i_3|\right] \\ &\quad + \ldots + \left(1 - p_{i_1} - \ldots - p_{i_{N-1}}\right)\left[t + |i_{N-1} - i_N|\right] . \end{aligned} \tag{2}$$

Each expression within the square brackets is minimized when $i_1, |i_1 - i_2|, \ldots, |i_{N-1} - i_N| = 1$. Also, each expression in the round brackets is minimized when $p_{i_1} \geq p_{i_2} \geq \ldots \geq p_{i_N}$. Since $p_1 \geq p_2 \geq \ldots \geq p_N$, and setting $(i_1, \ldots, i_N) = (1, 2, \ldots, N)$ makes $i_1, |i_1 - i_2|, \ldots, |i_{N-1} - i_N|$ attain their minima unity, hence the search order $(1, 2, \ldots, N)$ is best.

MODEL III — $p_i = 2i/N(N+1)$. For this model, exactly optimal strategies appear difficult to deduce. Of course, they could be found by enumerating the expected cost for each of the N! possible order of examination for each value of t; but this involves a computation of great magnitude. Furthermore, we may rationalize our way out of doing so by stating that, even if optimal, some of these policies may be eliminated from consideration as unrealistic in application: policies that demand that the search move backwards and forwards many times are unlikely to be very popular with the searcher.

Hence the policies to be considered from now on will be confined to the sets S, S' of the types shown in Figures 1 and 2. The policies are (R, R + 1, ... , N; R - 1, R - 2, ... , 2, 1) and (1, 2, ... , U; N, N - 1, ... , U + 1), respectively. That is, for set S go to R, then examine each cell from R to N, then go back to cell R - 1, and then search each cell from R - 1 to 1; for set S' examine cells 1 to U, then each cell from N to U + 1. These seem highly realistic kinds of search policies, and the objective of the analysis below is to determine the optimal values of R and U, and their relationship to N and t.

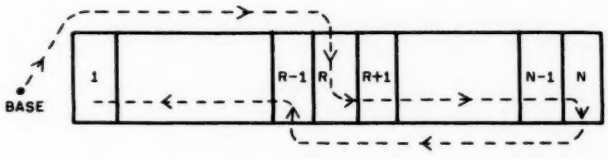

Figure 1

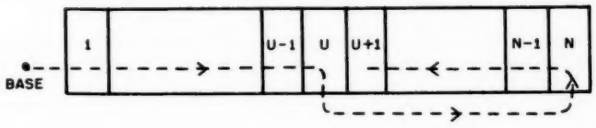

Figure 2

The Set S

The total cost T_i of the search if the object is in cell i is:

$$\underline{i < R}: \quad T_i = (R + t) + (N - R)(1 + t) + (N - R) + (R - i)(1 + t)$$

$$= N + t + (N - i)(1 + t) .$$

(3)

$$\underline{i = R}: \quad T_i = R + t .$$

$$\underline{i > R}: \quad T_i = (R + t) + (1 + t)(i - R)$$

$$= -t(R - 1) + i(1 + t) .$$

Then the expected cost $E(R)$ is given by

$$\frac{N(N+1)}{2} E(R) = \sum_{i=1}^{N} iT_i$$

$$= \frac{1}{2}(R-1)\,R[N+t+N(1+t)] - (1+t)\,[(R-1)\,R(R+1)/3 - R(R-1)/2] \tag{4}$$

$$+ (R+t)\,R - \frac{1}{2}t(R-1)\,[N(N+1) - R(R+1)]$$

$$+ (1+t)\,[N(N+1)\,(N+2)/3 - R(R+1)\,(R+2)/3 - N(N+1)/2 + R(R+1)/2] .$$

Defining $D(R) = [E(R+1) - E(R)]\,N(N+1)/2$, after some algebraic manipulation we obtain

$$D(R) = (N-R)\,[(t+4)\,R - (N+1)\,t]/2 . \tag{5}$$

Hence the minimum of the function $E(R)$ in the range $(1, N)$ is either at $R = N$ or at $R_1 = t(N+1)/(t+4)$. Now $R_1 < N-1$ if and only if $t < 2(N-1)$. It therefore follows that the optimal policies of the type under consideration are:

(i) $t \geq 2(N-1)$: Choose $R = N$. That is, examine in the order $(N, N-1, \ldots, 1)$.

(ii) $t < 2(N-1)$: Choose R equal to the nearest integer greater than R_1.

The Set S'

As in the previous sections, for the set S' we obtain

$$\underline{i < U}: \quad T: = i(1+t) .$$

$$\underline{i \geq U}: \quad T: = U(1+t) + (N-U+t) + (N-i)\,(1+t) \tag{6}$$

$$= t(N+U+1) + 2N - i(1+t) .$$

Then, as before, defining $D'(U) = [E(U+1) - E(U)]\,N(N+1)/2$, we obtain

$$D'(U) = -(N-U-1)\,[(t+4)\,(U+1) - (N+1)\,t]/2 . \tag{7}$$

It is of interest to note the similarity of $D'(U)$ and $D(R)$, and in fact we shall make use of the relationship

$$D'(V-1) = -D(V) \tag{8}$$

below.

Optimization Over S, S'

Consider the sequence of search patterns S(1), S(2), ... , S(N), S'(1), S'(2), ... , S'(N), where the numbers within parentheses refer to values of R and U. Firstly, we shall note that $\{E[S'(1)] - E[S(N)]\}\, N(N+1)/2 = -(N-1)[t + 4 - (N+1)t]$, which equals D'(0), and of course - D(1). Secondly, we note that $S(1) \equiv S'(N)$; i.e., the search patterns at the beginning and the end of the sequence are identical.

Now, for $t < \frac{4}{N}$, using equations (5), (7), and (8), we see that the expected cost increases monotonically from S(1) to S(N) and then decreases monotonically to S'(N), with the end values identical. This is because, for $t < \frac{4}{N}$, $D(R) > 0$ all R, and $D'(U) < 0$ all U. Hence S(1) (and S'(N)) give the minima of the expected cost function. That is, examination in the order (1, 2, ... , N) is optimal over the sets S, S'.

Similarly, for $t > 2(N-1)$, the expected cost decreases monotonically to S(N) and then increases monotonically to S'(N). Hence, examination in the order (N, N - 1, ... , 1) is optimal over S, S'.

Now, for the remaining interval $\frac{4}{N} \le t \le 2(N-1)$, the cost decreases monotonically to S(R), where R is the nearest integer greater than R_1, and then increases until D'(V) becomes negative, after which it decreases to S'(N). This follows from equation (8) from which it also follows that D'(V) becomes negative at $V = R_1 - 1$. It is therefore now clear that the optimal search pattern over S is the optimal pattern over the sum of S and S', since S contains the only minimum of the cost function. Hence, S(R), where R is the nearest integer greater than R_1, is the optimal search pattern.

Note that we have shown that S'(U) is never best. This is slightly comforting because there is an intuitive reason why policies in S are to be preferred: if the object is moving towards base, the only conditions under which it will be missed are if it is in one of cells 1 to R - 1 at the start and the searcher does not cover the return part of the search before the object reaches base. Since (a) the searcher will usually be traveling much faster than the object, and (b) the probability of the object being in a cell close to base is relatively small because of the model itself, this would imply a further justification of the set S.

Statistics of the Set S

It is of interest to note the dependence of R_1 on t and N, and to compare the cost of using the relatively optimal policy R_1 with that of using the policy that maximizes E(R) over the set S. Firstly, it should be noted that

$$E(1) = (1 + t)(2N + 1)/3 , \tag{9}$$

and

$$E(N) = (1 + t)(N - 1)/3 + N + t . \tag{10}$$

Hence, for all $t > 2$, $E(1) > E(N)$, and $E(1)$ is the maximum $E(R)$. Then, for $t \geq 2(N - 1)$,

$$K = \frac{\text{Max } E(R)}{\text{Min } E(R)} = \frac{E(1)}{E(N)} \sim 2, \tag{11}$$

so that the cost of the search is cut by a half approximately if we start at cell N rather than at cell 1. For $t < 2(N - 1)$, K is a much more complicated expression. Computations have been performed for $N = 5, 10, 20, 50, 100$, and $t = 2, 5, 10, 20, 30, \ldots, 4N$, and in Figure 3 K is plotted against t for these values of N.

Further, defining the value of R that minimizes $E(R)$ to be R_{min}, it should be noted that

$$R_{min}/N = t\left(1 + \frac{1}{N}\right)/(t + 4) \sim t/(t + 4) \quad \text{for large N.} \tag{12}$$

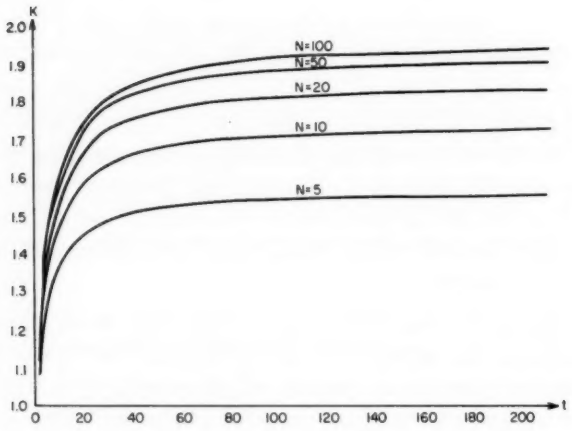

Figure 3

Summary of Policies for Model III

To summarize, the optimal "feasible" policies for Model III are:

$t \leq 4/N$:	Examine in the order $(1, 2, \ldots, N)$.
$4/N < t < 2(N - 1)$:	Examine in the order $(R, R + 1, \ldots, N, R - 1, R - 2, \ldots, 1)$, where R is the nearest integer greater than $R_1 = t(N + 1)/(t + 4)$.
$t \geq 2(N - 1)$:	Examine in the order $(N, N - 1, \ldots, 1)$.

REFERENCES

[1] R. Bellman, Dynamic Programming, Princeton University Press, 1957, page 90.

[2] W. E. Smith, "Various Optimizers for Single-Stage Production," Naval Research Logistics Quarterly, 3:62-63 (1956).

* * *

A DELIVERY-LAG INVENTORY MODEL WITH AN EMERGENCY PROVISION (THE SINGLE-PERIOD CASE)*†‡

E. W. Barankin

University of California

INTRODUCTION AND GENERAL DISCUSSION

This article initiates the study of an inventory model in which there is a fixed lag time of one period for delivery of orders, but in which there is also defined an emergency situation with respect to initial stock at any particular inventory point; when such an emergency situation obtains, an additional order of a specified fixed size is taken with immediate delivery. There is a certain additional cost for this immediate delivery (although the analysis may be specialized to the case of no additional cost). While the particular structure of emergency is laid down, there is a free parameter left to be chosen for each period to render the definition of emergency fully specific; and this parameter is determined as part of the optimization.

There is no attempt here to achieve the fullest generality in every respect. The principal object is to investigate the indicated emergency character of the model. It is anticipated that the n-period cases, with $n > 1$, will yield their essential properties by the device of inductive argument, as in [1], Chapt. 9, and that from these, by suitable limit considerations, the stationary case may be studied. Therefore, the main body of this first investigation is directed to the case of the single-period problem, so formulated as to present the features which one might hope to find in the general n-period case.

Before we enter into the single-period problem, it will be helpful for orientation to discuss the general n-period model, and the attendant recursion relations.

Consider a time interval of n periods in length, with the inventory points labelled $1, 2, \ldots, n+1$ from left to right, and the periods labelled I_i, $i = 1, 2, \ldots, n$, as shown in Figure 1. For $i = 1, 2, \ldots, n$, let x_i denote what shall be called the initial stock in I_i, and is defined to be the (positive, zero or negative) stock level at i inclusive of any orders that have been placed at inventory points $t < i$ and which arrive at i, but not inclusive of an order

*Manuscript received April 1961.

†This work received partial support of the Office of Naval Research (Nonr-222-43). Reproduction in whole or in part is permitted for any purpose of the United States Government.

‡This article is a revision of the Report No. 17 submitted to the Office of Naval Research in the summer of 1960. The corrections and revisions made (in particular, the introduction of the credit function v into the discussion of the n-period case in the first section, and the correction of the Lemma in the next-to-the-last section) are due largely to the diligent study of Mr. K. Daniel.

The original version of this article constitutes part of the proceedings of the Decennial Logistics Research Conference held at George Washington University in 1960. Other papers from this Conference have been published in the December 1960 and 1961 issues of the Naval Research Logistics Quarterly.

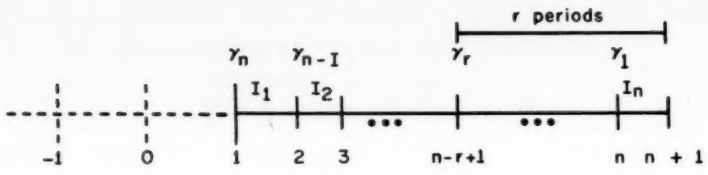

Figure 1

which is placed at i and is immediately delivered. For purposes of deriving the recursion relation, it is necessary to consider the n-period interval in the context of a larger interval (as indicated in Figure 1, with inventory points 0, -1, ...); thus, the initial stock x_1 takes account of orders at t = 1 which were placed at previous inventory points.

Now in fact, we consider that an order arrives at an inventory point only on either immediate delivery or delivery of one period lag. In more detail: at a point i, ordering is done according to the following scheme. There is a number γ_{n-i+1} such that (i) if $x_i > \gamma_{n-i+1}$, then an order for some number z_i of units (the quantity z_i to be determined by optimization) is placed for delivery at i + 1, one period later, whereas (ii) if $x_i \leqq \gamma_{n-i+1}$, then in addition to placing an optimally determined order to be delivered at i + 1, a further order for m units is placed and these m units will be immediately delivered. The quantity m is fixed, the same for all periods. The numbers γ_j are to be chosen in an optimal way.

The sense of this ordering scheme is clear. We are taking the simplest kind of characterization of an emergency situation; namely, when the initial stock is below a prescribed level (the number γ_j). And in this situation we are considering that the supplier will accommodate with an immediate delivery, though not of an arbitrarily large order, but, more realistically, of a fixed quantity m which he is able to produce with relative ease (but at perhaps somewhat higher cost—see below) for immediate delivery. The inventory manager is free to set the values of the γ_j, thus fully defining the states of stock-level emergency. We shall see that in certain cases there are simple optimal selections of the γ_j.

We shall employ the term starting stock at the inventory point i to designate (i) the initial stock itself if there is no ordering at i of the additional m units for immediate delivery (no emergency), or (ii) the initial stock plus m if there is ordering of the additional m units for immedate delivery (emergency).

Let the demands in the several inventory periods be independent and identically distributed, with distribution having a continuous density φ:

$$\varphi(\xi)\begin{cases} = 0 \text{ for } \xi < 0 \\ \geqq 0 \text{ and continuous for } \xi > 0\,. \end{cases} \tag{1}$$

We assume that $\varphi(0+)$ exists, and we shall denote this limit more simply by $\varphi(0)$.

We consider a holding cost function, h, and a penalty cost function, p, applicable in each period. For any particular period, if y is the starting stock level at the beginning of this period and ξ is the demand in this period, then the holding cost is $h(y - \xi)$ and the penalty cost is $p(\xi - y)$, these costs being charged at the end of the period. Of course, only one of

these two costs may be positive in any particular case, as follows from the first assumptions on h and p:

$$h(\eta)\begin{cases} = 0 \quad \underline{\text{for}} \quad \eta \leqq 0 \\ \geqq 0 \quad \underline{\text{for}} \quad \eta > 0 \\ \text{convex for all } \eta\,. \end{cases} \tag{1.2}$$

$$p(\eta)\begin{cases} = 0 \quad \underline{\text{for}} \quad \eta \leqq 0 \\ \geqq 0 \quad \underline{\text{for}} \quad \eta > 0 \\ \text{convex for all } \eta\,. \end{cases} \tag{1.3}$$

If $y - \xi$ is nonnegative for a certain period I_i, then the costing for that period is completed with the calculation of $h(y - \xi)$ at the end of the period. However, if $y - \xi$ is negative, then we consider that the net penalty cost for the period is not necessarily the full penalty charge, $p(\xi - y)$, but rather is the resultant of this charge and a credit for as many of the $\xi - y$ unfilled demands as can be met at the end of the period, that is, at $t = i + 1$, with the incoming regular delivery of z units that were ordered at $t = i$. Let v denote the credit function thus assumed. The arriving z units can fill $\min[z, \xi - y]$ of the $\xi - y$ back-logged demands, so that the net penalty cost for the period is

$$p(\xi - y) - v(\min[z, \xi - y])\,. \tag{1.4}$$

The filling of some or all of the $\xi - y$ demands at the end of the period with the arriving z units is a late accommodation of these demands, and therefore we consider that credits cannot in general fully compensate penalties; and in fact, we suppose the function v to satisfy:

$$v(\eta)\begin{cases} = 0 \quad \underline{\text{for}} \quad \eta \leqq 0\,, \\ \geqq 0 \quad \underline{\text{for}} \quad \eta > 0\,, \\ \text{concave for } \eta > 0\,. \end{cases} \tag{1.5}$$

We take a fixed cost c per unit of the item under consideration. However, this is the regular cost which is applicable only in the case of orders to be regularly delivered, that is, to be delivered one period later. In the case of an emergency, the m units that are delivered immediately will be considered to have a unit cost c_0, not less than c and in general greater than c. Thus, in an emergency, the m immediately delivered units will have the cost $c_0 m$, while the cost of the simultaneously placed, regular order, z, will be cz (the latter cost being incurred at the time of placement of the order).

We assume that there is no set-up cost. The discount factor will be denoted by α.

617404 O - 61 - 6

Now, for $r = 1, 2, \ldots, n$, let

(1.6) $\bar{f}_r(x, z, \xi; \gamma_r)$ = conditional expected total cost, discounted to $t = n - r + 1$, for the r-period interval from $t = n - r + 1$ to $t = n + 1$, given that the initial stock at $n - r + 1$ is x, that a regular order for z units is placed at $n - r + 1$, and that the demand in the period $[n - r + 1, n - r + 2]$ is ξ; and given, also, that there is an optimal ordering rule in effect in $[n - r + 2, n + 1]$, and that γ_r specifies the emergency level at $n - r + 1$.

Let

$$(1.7) \qquad f_r^0(x, z; \gamma_r) = \int_{\xi=0}^{\infty} \bar{f}_r(x, z, \xi; \gamma_r)\, \varphi(\xi)\, d\xi \ .$$

And, finally, let

(1.8) $f_r(x)$ = optimal value of $f_r^0(x, z; \gamma_r)$ among all choices of

(i) the constant γ_r independent of x, and

(ii) the quantity z dependent upon x and the choice of γ_r.

We have purposely not used the word "minimal" in (1.8) where the word "optimal" appears, because it is not in general true that there is a unique minimum for $f_r^0(x, z; \gamma_r)$ under the condition (i) in (1.8) which demands that γ_r be fixed independently of x. To formulate this more precisely, let

$$(1.9) \qquad f_r^{(1)}(x; \gamma_r) = \min_{z \geq 0} f_r^0(x, z; \gamma_r)$$

and

$$(1.10) \qquad f_r^{(2)}(x) = \min_{\gamma_r} f_r^{(1)}(x; \gamma_r) \ .$$

Then, it makes sense to substitute the word "minimal" in (1.8) only if there is some value of γ_r, say γ_r', such that

$$(1.11) \qquad f_r^{(2)}(x) = f_r^{(1)}(x; \gamma_r'), \text{ all } x \ ;$$

and in this case the value γ_r' will be taken to define the emergency level, and the optimal cost function f_r of (1.8) will be $f_r^{(2)}$. If there is no such value γ_r', then some new convention must

be made as to what "optimal" shall mean. For each r, we consider that such a new convention is at hand, if necessary, to provide the Functions f_r of (1.8).

Now, let us examine this question of minimality more closely by looking at f_r^o as given by the recursion relation. For $r > 1$ we have

$$\bar{f}_r(x,z,\xi;\gamma_r) = \begin{cases} cz + l(x-\xi) - v(\min[z, \xi - x]) \\ \quad + \alpha f_{r-1}(x+z-\xi), \qquad x > \gamma_r, \\ cz + c_o m + l(x+m-\xi) - v(\min[z, \xi - (x+m)]) \\ \quad + \alpha f_{r-1}(x+m+z-\xi), \quad x \leqq \gamma_r, \end{cases} \tag{1.12}$$

where l is the function defined by

$$l(\eta) = \begin{cases} h(\eta), & \eta > 0, \\ p(-\eta), & \eta \leqq 0, \end{cases} \tag{1.13}$$

If we set

$$L(\eta) = \begin{cases} \int_{\xi=0}^{\eta} h(\eta-\xi)\,\varphi(\xi)\,d\xi + \int_{\xi=\eta}^{\infty} p(\xi-\eta)\,\varphi(\xi)\,d\xi, & \eta > 0, \\ \int_{\xi=0}^{\infty} p(\xi-\eta)\,\varphi(\xi)\,d\xi, & \eta \leqq 0, \end{cases} \tag{1.14}$$

and

$$V(x,z) = \begin{cases} \int_{\xi=0}^{x+z} v(\xi-x)\,\varphi(\xi)\,d\xi + v(z) \int_{\xi=x+z}^{\infty} \varphi(\xi)\,d\xi, & x+z > 0, \\ v(z), & x+z \leqq 0, \end{cases} \tag{1.15}$$

then we have, from (1.12),

$$f_r^o(x,z;\gamma_r) = \begin{cases} cz + L(x) - V(x,z) \\ \quad + \alpha \int_{\xi=0}^{\infty} f_{r-1}(x+z-\xi)\,\varphi(\xi)\,d\xi, \quad x > \gamma_r, \\ cz + c_o m + L(x+m) - V(x+m,z) \\ \quad + \alpha \int_{\xi=0}^{\infty} f_{r-1}(x+m+z-\xi)\,\varphi(\xi)\,d\xi, \quad x \leqq \gamma_r. \end{cases} \tag{1.16}$$

Let us set, for $r \geqq 1$:

$$G_r(z;\eta) = cz - V(\eta,z) + \alpha \int_{\xi=0}^{\infty} f_r(\eta + z - \xi)\,\varphi(\xi)\,d\xi \,. \tag{1.17}$$

Then (1.16) becomes

$$f_r^0(x,z;\gamma_r) = \begin{cases} L(x) + G_{r-1}(z;x)\,, & x > \gamma_r\,, \\ c_0 m + L(x+m) + G_{r-1}(z;x+m), & x \leqq \gamma_r\,. \end{cases} \tag{1.18}$$

Let us put

$$H_r(\eta) = \min_{z \geqq 0} G_{r-1}(z;\eta)\,, \tag{1.19}$$

and

$$J_r(\eta) = L(\eta) + H_r(\eta)\,. \tag{1.20}$$

Then we have, according to (1.9),

$$f_r^{(1)}(x;\gamma_r) = \begin{cases} J_r(x)\,, & x > \gamma_r\,, \\ c_0 m + J_r(x+m)\,, & x \leqq \gamma_r\,. \end{cases} \tag{1.21}$$

We see thus that $f_r^{(1)}$ has a rather special form, and we can inquire after conditions under which (1.11) holds.

From (1.21) we find that

$$f_r^{(2)}(x) = \min\left[J_r(x)\,,\ c_0 m + J_r(x+m)\right]. \tag{1.22}$$

Hence, if there is a γ_r' such that (1.11) holds, we have

$$\begin{cases} J_r(x) \leqq c_0 m + J_r(x+m)\,, & x > \gamma_r'\,, \\ J_r(x) \geqq c_0 m + J_r(x+m)\,, & x \leqq \gamma_r'\,. \end{cases} \tag{1.23}$$

Conversely, if there is a γ_r' satisfying (1.23), then (1.11) holds. Thus, the existence of a γ_r' such that (1.23) holds is, in our case, the characteristic condition on the function J_r in order that optimality in (1.8) be definable as minimality.

Notice that the function $c_0 m + J_r(x+m)$ is, geometrically, the translate of the function $J_r(x)$ first to the left by m and then up by the amount $c_0 m$. If the Function J_r is U-shaped,

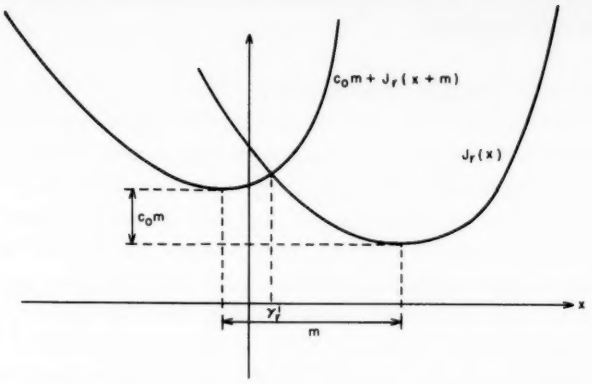

Figure 2

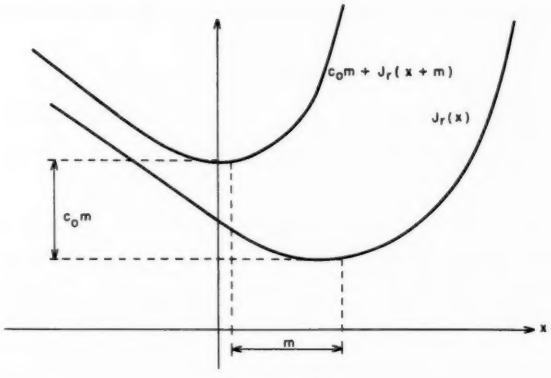

Figure 3

the picture may be as shown in Figure 2. In this case—that is, a situation of the type depicted in Figure 2—the γ_r' indicated in the figure is the unique number verifying (1.23).

But observe that—the nature of the Function J_r permitting—c_0 might be so large that, instead of the situation in Figure 2, we have what is shown in Figure 3. This is the situation that

$$J_r(x) \leqq c_0 m + J_r(x+m), \quad \text{all } x. \tag{1.24}$$

In this case, we may say that (1.23) holds with $\gamma_r' = -\infty$. And, operationally, this means that we shall not admit any emergency situation—a natural consequence of too large an emergency cost as compared with penalties.

In subsequent investigations, we expect to study the question of convexity of the successive J_r's, and the existence of finite values γ_r', $r = 1, 2, \ldots$ as shown in Figure 2. For the present, we direct the necessary first attention to the single-period case, and find that

under natural assumptions we do have the indicated convexity of J_1 and consequent analysis according to Figure 2 or Figure 3.

FORMULATION OF THE SINGLE-PERIOD CASE

As we have already said, our intention in fully formulating and analyzing the one-period case is to accomplish the prototype of argumentation that we would hope to be applicable in the general n-th stage just discussed. We shall therefore consider that there is an initial stock level of x at the beginning of our single period, that there is an emergency level γ, as previously described, to be determined (in particular, it might turn out that we would want to take $\gamma = -\infty$ (disallowance of any emergency situation)), and that a regular order for z units is placed at the beginning of the period for delivery at the end of the period. To complete the description, we consider that the z units arriving at the end of the period may be used to fill any demands still outstanding. For this late filling of demands the credit function v is applicable as described in the first section. Finally, we assume that if, after late filling of demands, there is still some stock left over, it is sold for salvage, the salvage gain function being w.

The functions v and w will be supposed concave for positive arguments; we assume, in fact, the following properties of v and w and their first and second derivatives:

$$(2.1)\qquad \begin{cases} v(\eta) = 0, & \eta \leqq 0, \\ v'(\eta) > 0, & \eta \geqq 0 \quad (v'(0) \text{ denotes } v'(0+)), \\ v''(\eta) \leqq 0, & \eta \geqq 0 \quad (v''(0) \text{ denotes } v''(0+)). \end{cases}$$

$$(2.2)\qquad \begin{cases} w(\eta) = 0, & \eta \leqq 0, \\ w'(\eta) > 0, & \eta \geqq 0 \quad (w'(0) \text{ denotes } w'(0+)), \\ w''(\eta) \leqq 0, & \eta \geqq 0 \quad (w''(0) \text{ denotes } w''(0+)). \end{cases}$$

Correspondingly, we make the following more detailed assumptions concerning the holding and penalty functions h and p:

$$(2.3)\qquad \begin{cases} h(\eta) = 0, & \eta \leqq 0 \\ h'(\eta) > 0, & \eta \geqq 0 \quad (h'(0) \text{ denotes } h'(0+)), \\ h''(\eta) \geqq 0, & \eta \geqq 0 \quad (h''(0) \text{ denotes } h''(0+)). \end{cases}$$

$$(2.4)\qquad \begin{cases} p(\eta) = 0, & \eta \leqq 0 \\ p'(\eta) > 0, & \eta \geqq 0 \quad (p'(0) \text{ denotes } p'(0+)), \\ p''(\eta) \geqq 0, & \eta \geqq 0 \quad (p''(0) \text{ denotes } p''(0+)). \end{cases}$$

Our analysis will be carried out under several assumptions regarding the interrelations between the Functions v, w, h, and p, and the cost Constant c. These are the following:

(2.5) $$v'_\infty > w'(0), \qquad (v'_\infty \equiv \lim_{\eta \to \infty} v'(\eta)),$$

(2.6) $$v'_\infty > c > w'_\infty, \qquad (w'_\infty \equiv \lim_{\eta \to \infty} w'(\eta)),$$

(2.7) $$v'(0) < p'(0),$$

and

(2.8) $$h'_\infty > w'_\infty, \qquad (h'_\infty \equiv \lim_{\eta \to \infty} h'(\eta)).$$

The significance of Condition (2.5) is that there is always a greater gain in filling demands late, no matter how large the amount of such demands, than in leaving some of these demands unfulfilled and taking the salvage value of the left-over stock instead. This interpretation makes use, of course, of the fact that the second derivatives v'' and w'' are nonpositive for all nonnegative arguments, so that we have, in fact,

(2.9) $$v'(\eta_1) \geqq v'_\infty > w'(0) \geqq w'(\eta_2) \quad \text{for all } \eta_1 \geqq 0 \text{ and } \eta_2 \geqq 0.$$

The left-hand inequality in (2.6) means—again taking account of the fact that $v'(\eta) \geqq v'_\infty$ for all $\eta \geqq 0$—that every demand filled late, no matter how large the amount of such demands, represents a gain in excess of the (regular) purchase cost of the quantity of item needed to fill that demand. In other words, even late filling of demands, however many, has sufficient worth that the purchase cost of the item is no deterrent.

Our assumptions on the function w imply, of course, that w' is monotonely nonincreasing from 0 to ∞. In view of this, the right-hand inequality in (2.6) means that there is a level beyond which salvage returns do not make up purchase cost. Indeed, if the reverse were true there would be gain to be had in the purely subsidiary operation of purchasing for resale to salvage.

The relation (2.7)—again together with the consequences of the second derivative conditions on v and p—asserts the realistic situation that the gain achieved by filling demands late does not make up all of the penalty that is paid for this lateness.

And finally, taking account of the second derivative conditions on w and h, Conditions (2.8) means that there is no level of stock sufficiently large that beyond this level salvage returns make up holding costs.

Let γ denote the (to be optimally determined) single-period emergency level, and let $\bar{f}$ stand for $\bar{f}_1$. Then we have

$$\bar{f}(x,z,\xi;\gamma) = \begin{cases} cz + 1(x-\xi) - v(\min[z,\ \xi - x]) \\ \qquad - w(x+z-\xi)\,, & x > \gamma\,, \\ cz + c_0 m + 1(x+m-\xi) \\ \qquad - v(\min[z,\ \xi - (x+m)]) \\ \qquad - w(x+m+z-\xi)\,, & x \leqq \gamma\,. \end{cases} \tag{2.10}$$

Letting f^0 stand for f_1^0, we thus have

$$f^0(x,z;\gamma) = \begin{cases} cz + L(x) - V(x,z) - W(x+z)\,, & x > \gamma\,, \\ cz + c_0 m + L(x+m) - V(x+m,\,z) \\ \qquad - W(x+m+z)\,, & x \leqq \gamma\,, \end{cases} \tag{2.11}$$

where

$$V(x,z) = \begin{cases} \displaystyle\int_0^{x+z} v(\xi - x)\,\varphi(\xi)\,d\xi + v(z) \int_{x+z}^{\infty} \varphi(\xi)\,d\xi\,, \\ \qquad\qquad x + z > 0\,, \\ v(z)\,, \qquad x + z \leqq 0\,, \end{cases} \tag{2.12}$$

and

$$W(x) = \int_0^{\infty} w(x-\xi)\,\varphi(\xi)\,d\xi\,. \tag{2.13}$$

Define

$$G(z;x) = cz - V(x,z) - W(x+z)\,; \tag{2.14}$$

then, (2.11) becomes

$$f^0(x,z;\gamma) = \begin{cases} L(x) + G(z;x)\,, & x > \gamma\,, \\ c_0 m + L(x+m) + G(z;\,x+m)\,, & x \leqq \gamma\,. \end{cases} \tag{2.15}$$

We put

$$H(x) = \min_{z \geqq 0} G(z;x) \tag{2.16}$$

and

$$J(x) = L(x) + H(x) . \tag{2.17}$$

Then we have, letting $f^{(1)}$ stand for $f_1^{(1)}$,

$$f^{(1)}(x;\gamma) = \begin{cases} J(x) , & x > \gamma , \\ c_0 m + J(x+m) , & x \leqq \gamma . \end{cases} \tag{2.18}$$

We see that $f^{(1)}$ is of the same form as $f_r^{(1)}$ for $r > 1$, as given by (1.21). Hence, the deliberations in the first section subsequent upon formula (1.21) are applicable here in the single-period case (our present G, H, and J being G_0, H_1, and J_1, resp., in the notation of the first section) and the problem facing us is the study of the function J with the goal of determining whether or not there exists a γ' such that

$$\begin{cases} J(x) \leqq c_0 m + J(x+m) , & x > \gamma' , \\ J(x) \geqq c_0 m + J(x+m) , & x \leqq \gamma' . \end{cases} \tag{2.19}$$

STUDY OF THE MINIMIZATION OF $G(z;x)$

Our first major task toward understanding the behavior of the function J is to get at the function H; that is, to study the minimization of $G(z;x)$ with respect to $z \geqq 0$.

We notice, to begin with, that for each x, $V(x,z)$ and $W(x+z)$ are continuous functions of $z \geqq 0$. Hence, $G(z;x)$ is, for each x, a continuous function of z. In fact, G is differentiable with respect to z, and to calculate this partial derivative and other derivatives further along in our analysis, we set down the following more explicit (than (2.12) and (2.13)) formulas:

$$V(x,z) = \begin{cases} \displaystyle\int_0^{x+z} v(\xi - x)\,\varphi(\xi)\,d\xi + v(z)\int_{x+z}^{\infty} \varphi(\xi)\,d\xi , \\ \qquad x+z > 0 , \quad x \leqq 0 , \\ \displaystyle\int_x^{x+z} v(\xi - x)\,\varphi(\xi)\,d\xi + v(z)\int_{x+z}^{\infty} \varphi(\xi)\,d\xi , \\ \qquad x+z > 0 , \quad x > 0 , \\ v(z) , \quad x+z \leqq 0 , \end{cases} \tag{3.1}$$

$$W(x+z) = \begin{cases} \displaystyle\int_0^{x+z} w(x+z-\xi)\,\varphi(\xi)\,d\xi , & x+z > 0 , \\ 0 , & x+z \leqq 0 . \end{cases} \tag{3.2}$$

From (3.1)—or, in fact, also from (2.12)—we have

$$\frac{\partial V}{\partial z} = \begin{cases} v'(z) \int_{x+z}^{\infty} \varphi(\xi)\, d\xi\,, & x + z > 0\,, \\ v'(z)\,, & x + z \leqq 0\,; \end{cases} \tag{3.3}$$

and from (3.2) we get (on taking account of the fact that $w(0) = 0$)

$$\frac{\partial W(x+z)}{\partial z} = \begin{cases} \int_{0}^{x+z} w'(x+z-\xi)\, \varphi(\xi)\, d\xi\,, & x + z > 0\,, \\ 0\,, & x + z \leqq 0\,. \end{cases} \tag{3.4}$$

Hence, we have

$$\frac{\partial G}{\partial z} = \begin{cases} c - v'(z) \int_{x+z}^{\infty} \varphi(\xi)\, d\xi - \int_{0}^{x+z} w'(x+z-\xi)\, \varphi(\xi)\, d\xi\,, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x + z > 0\,, \\ c - v'(z)\,, \quad x + z \leqq 0\,. \end{cases} \tag{3.5}$$

We see readily that, for each x, $\partial G/\partial z$ is continuous in $z > 0$. In fact, $\partial G/\partial z$ is differentiable with respect to z; but before we go ahead to find $\partial^2 G/\partial z^2$, we first examine the behavior of $\partial G/\partial z$.

From (3.5) we find

$$\left(\frac{\partial G}{\partial z}\right)_{z=0+} = \begin{cases} c - v'(0) \int_{x}^{\infty} \varphi(\xi)\, d\xi - \int_{0}^{x} w'(x-\xi)\, \varphi(\xi)\, d\xi\,, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x > 0\,, \\ c - v'(0)\,, \quad x \leqq 0\,. \end{cases} \tag{3.6}$$

By (2.1) and (2.6) we have

$$v'(0) \geqq v'_{\infty} > c\,. \tag{3.7}$$

Applying this to (3.6) we see that

$$\left(\frac{\partial G}{\partial z}\right)_{z=0+} < 0 \quad \text{for each } x \leqq 0\,. \tag{3.8}$$

The function $(\partial G/\partial z)_{z=0+}$ is a continuous function of x, constant and negative for $x \leqq 0$. Let us examine its derivative for $x > 0$:

$$\frac{d}{dx}\left[\left(\frac{\partial G}{\partial z}\right)_{z=0+}\right] = v'(0)\,\varphi(x) - w'(0)\,\varphi(x) - \int_0^x w''(x-\xi)\,\varphi(\xi)\,d\xi\,, \quad x > 0\,. \tag{3.9}$$

By (2.1) and (2.5),

$$v'(0) \geqq v'_\infty > w'(0)\,. \tag{3.10}$$

Combining this with the fact that $\varphi(x) \geqq 0$, and remembering that $w'' \leqq 0$, we see that

$$\frac{d}{dx}\left[\left(\frac{\partial G}{\partial z}\right)_{z=0+}\right] \geqq 0\,, \quad x > 0\,. \tag{3.11}$$

Hence, $(\partial G/\partial z)_{z=0+}$ is monotone nondecreasing in x. We find, furthermore, from (3.6),

$$\lim_{x\to\infty}\left(\frac{\partial G}{\partial z}\right)_{z=0+} = c - w'_\infty > 0 \tag{3.12}$$

as a consequence of (2.6). It follows that $(\partial G/\partial z)_{z=0+}$ vanishes for some unique positive value of x or for some closed interval of positive values of x. Let x_o be this unique value of x, if it exists, or the lower endpoint of this interval of values of x, if an interval exists. Then, x_o is either the unique or the minimum number verifying

$$c - v'(0)\int_{x_o}^{\infty}\varphi(\xi)\,d\xi - \int_0^{x_o} w'(x_o-\xi)\,\varphi(\xi)\,d\xi = 0\,. \tag{3.13}$$

To sum up what we have found thus far:

(3.14)

There is a positive number x_o such that the continuous, monotone nondecreasing function (3.6) satisfies

$$\left(\frac{\partial G}{\partial z}\right)_{z=0+} \begin{cases} < 0\,, & x < x_o\,, \\ \geqq 0\,, & x \geqq x_o\,. \end{cases}$$

The number x_o is the smallest solution of (3.13).

Let us observe that if φ is strictly positive for positive arguments then the inequality sign holds in (2.11). And as a consequence of this the function (3.6) cannot vanish on an interval. Hence:

(3.15) If $\varphi(\eta) > 0$ for $\eta > 0$, then the number x_0 of (3.14) is the unique solution of (3.13), and is consequently the only point at which the function (3.6) vanishes.

Let us now study the second derivative (obtained from (3.5)):

$$\frac{\partial^2 G}{\partial z^2} = \begin{cases} -v''(z) \int_{x+z}^{\infty} \varphi(\xi)\, d\xi - \int_0^{x+z} w''(x+z-\xi)\, \varphi(\xi)\, d\xi \\ \qquad + [v'(z) - w'(0)]\, \varphi(x+z), \\ \qquad\qquad x + z > 0, \\ -v''(z), \quad x + z < 0. \end{cases} \tag{3.16}$$

Since v'' and w'' are nonpositive, and φ is nonnegative and $v'(z) > w'(0)$ (see (2.9)), we see by (3.16) that $\partial^2 G/\partial z^2$ is, for each x, nonnegative for all $z \geqq 0$ for which it is defined, namely, (i) for all nonnegative z in case $x \geqq 0$, and (ii) for all nonnegative z except possibly $z = -x$ in case $x < 0$. In either case the derivative $\partial G/\partial z$ is therefore monotone nondecreasing in z. For $x < 0$ we have

$$\lim_{z \downarrow -x} \frac{\partial^2 G}{\partial z^2} = -v''(-x), \quad x < 0, \tag{3.17}$$

and

$$\lim_{z \downarrow -x} \frac{\partial^2 G}{\partial z^2} = -v''(-x) + [v'(-x) - w'(0)]\, \varphi(0), \quad x < 0. \tag{3.18}$$

Thus, by (2.9),

$$\lim_{z \downarrow -x} \frac{\partial^2 G}{\partial z^2} \geqq \lim_{z \downarrow -x} \frac{\partial^2 G}{\partial z^2}, \quad x < 0. \tag{3.19}$$

Strict inequality holds here if $\varphi(0) > 0$, as we see by (2.9) and (3.18). We have, as noted, the result that $\partial G/\partial z$ is monotone nondecreasing in z, for each x. For $x \geqq x_0$ (see (3.14)), this means immediately that $G(z, x)$ has a minimum with respect to z and it takes on this minimum value at $z = 0$. In particular, for an $x > x_0$, such that $(\partial G/\partial z)_{z=0+} > 0$, the minimum

of G is taken on only at $z = 0$. Hence, under the condition of (3.15), $z = 0$ is the unique minimum point of G for each $x > x_0$. For $x = x_0$, we see by (3.16) that

$$\left[\left(\frac{\partial^2 G}{\partial z^2}\right)_{z=0+}\right]_{x=x_0} = -v''(0)\int_{x_0}^{\infty}\varphi(\xi)\,d\xi$$

(3.20)

$$-\int_0^{x_0} w''(x_0 - \xi)\,\varphi(\xi)\,d\xi + [v'(0) - w'(0)]\,\varphi(x_0),$$

and this is strictly positive under the condition that $\varphi(\eta) > 0$ for $\eta > 0$. Hence, under the condition of (3.15), it is true also for $x = x_0$ that the minimum of G is taken on only at $z = 0$.

We must now examine the minimization of G, with respect to z, for values of $x < x_0$. For such a value of x, $G(z;x)$ starts out at $z = 0$ with a negative slope (see (3.14)). This slope is monotone nondecreasing, by virtue of the nonnegativity of $\partial^2 G/\partial z^2$. We ask now if this slope eventually becomes positive. By (3.5) we have

(3.21)
$$\lim_{z\to\infty}\frac{\partial G}{\partial z} = c - w'_{\infty} > 0.$$

Thus, the answer to our question is in the affirmative; in fact, $\partial G/\partial z$ tends to the same positive limit in $z \to \infty$ for every x. Therefore, in the case of concern to us, namely, $x < x_0$, the derivative $\partial G/\partial z$ vanishes at some unique positive value of z or for all values of z in some positive interval, and at this unique point, or at all the points of this interval, as the case may be, the function G takes on its minimum value in z. Now, by (2.6) and the monotone nonincreasing character of v', we see in (3.5) that $\partial G/\partial z < 0$ for every z such that $x + z \leqq 0$. It follows that the minimizing value or values of z must be such that $x + z > 0$. Then, by (3.16) we see that under the condition of (3.15), namely, $\varphi(\eta) > 0$ for $\eta > 0$, the second derivative $\partial^2 G/\partial z^2$ is strictly positive at a minimizing z. Consequently, there is a unique minimizing z under this condition. And this minimizing z is the value that causes the vanishing of the first expression in (3.5).

Let us now gather together our results on the minimization of G:

Proposition I

For each $x \geqq x_0$, $G(z;x)$ is monotone, nondecreasing, convex in $z \geqq 0$. Its minimum is therefore taken on either at $z = 0$ only or at all points of a closed interval with left endpoint $z = 0$. If $\varphi(\eta) > 0$ for $\eta > 0$, then $z = 0$ is the unique minimum point.

For each $x < x_0$, $G(z;x)$ is first decreasing and then increasing as z runs from 0 to ∞, and it is convex. Its minimum is taken on either at a unique positive point or at every point of a closed positive interval. Such a minimum point z satisfies $x + z > 0$ and is a root of the equation

(3.22)
$$c - v'(z)\int_{x+z}^{\infty}\varphi(\xi)\,d\xi - \int_0^{x+z} w'(x + z - \xi)\,\varphi(\xi)\,d\xi = 0.$$

If $\varphi(\eta) > 0$ for $\eta > 0$, then there is a unique minimum point for $G(z;x)$, and this point is the unique solution of (3.22).

From this stage on we shall work under the following assumption, which has been shown above to be important for uniqueness of the policy function:

ASSUMPTION:

$$\varphi(\eta) > 0 \quad \underline{\text{for}} \ \eta > 0 . \tag{3.23}$$

With this assumption we may now define

$$z^*(x) \stackrel{\text{def}}{=\!=} \text{the unique value of } z \text{ at which } G(z,x) \text{ takes on its minimum value.} \tag{3.24}$$

From Proposition I we have:

$$z^*(x) = \begin{cases} 0 , & \underline{\text{for}} \ x \geq x_o , \\ \text{the unique solution of (3.22), for } x < x_o , \end{cases} \tag{3.25}$$

and the additional fact that

$$x + z^*(x) > 0 \quad \text{for all } x . \tag{3.26}$$

STUDY OF THE FUNCTION $H(x)$

We have

$$H(x) = \min_{z \geq 0} G(z;x) = G(z^*(x) ; x) , \tag{4.1}$$

and we shall examine the first and second derivatives of H.

In general,

$$H'(x) = \left(\frac{\partial G}{\partial x}\right)_{z=z^*(x)} + \left(\frac{\partial G}{\partial z}\right)_{z=z^*(x)} \frac{dz^*}{dx} \tag{4.2}$$

$$\text{for } x \neq x_o \text{ (possibly)} .$$

But we have

$$\begin{cases} \left(\dfrac{\partial G}{\partial z}\right)_{z=z^*(x)} = 0 \quad \text{for } x < x_o , \\ \dfrac{dz^*}{dx} = 0 \quad \text{for } x > x_o . \end{cases} \tag{4.3}$$

Hence, (4.2) reduces to

$$H'(x) = \left(\frac{\partial G}{\partial x}\right)_{z=z^*(x)}, \quad x \neq x_0. \tag{4.4}$$

From (2.14),

$$\frac{\partial G}{\partial x} = -\frac{\partial V}{\partial x} - \frac{\partial W(x+z)}{\partial x}. \tag{4.5}$$

By virtue of (3.26) we see that the first two expressions in (3.1) and the first expression in (3.2) are the pertinent ones for evaluation of the terms on the right-hand side of (4.5). We find

$$\frac{\partial G}{\partial x} = \begin{cases} \displaystyle\int_0^{x+z} v'(\xi - x)\,\varphi(\xi)\,d\xi - \int_0^{x+z} w'(x+z-\xi)\,\varphi(\xi)\,d\xi, & \\ & x \leqq 0, \\ \displaystyle\int_x^{x+z} v'(\xi - x)\,\varphi(\xi)\,d\xi - \int_0^{x+z} w'(x+z-\xi)\,\varphi(\xi)\,d\xi, & \\ & x > 0. \end{cases} \tag{4.6}$$

for $x+z>0$

From (3.13) and (3.22), it follows that z^* is a continuous function, and this together with the continuity properties of $\partial G/\partial x$ in (4.6) implies that H' is a continuous function for all x. Thus, the exception noted in (4.4), namely, $x \neq x_0$, may be discarded.

We may obtain "explicit" expressions for $H'(x)$ by substituting $z^*(x)$ for z in (4.6). If we do this and avail ourselves of (3.22), we obtain:

$$H'(x) = \begin{cases} \displaystyle\int_0^{x+z^*(x)} v'(\xi - x)\,\varphi(\xi)\,d\xi + v'(z^*(x)) \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi - c, & \\ & x \leqq 0, \\ \displaystyle\int_x^{x+z^*(x)} v'(\xi - x)\,\varphi(\xi)\,d\xi + v'(z^*(x)) \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi - c, & \\ & 0 < x < x_0, \\ \displaystyle -\int_0^{x} w'(x-\xi)\,\varphi(\xi)\,d\xi, \quad x \geqq x_0. & \end{cases} \tag{4.7}$$

From the third expression here, we get immediately

$$\lim_{x \to \infty} H'(x) = - w'_{\infty} . \tag{4.8}$$

To gain information about the behavior of H' as $x \to - \infty$, we must study z^* a little more closely. From (3.22) we find:

$$\frac{dz^*}{dx} = - \frac{[v'(z^*(x)) - w'(0)]\, \varphi(x + z^*(x)) - \int_0^{x+z^*(x)} w''(x + z^*(x) - \xi)\, \varphi(\xi)\, d\xi}{\left\{ [v'(z^*(x)) - w'(0)]\, \varphi(x + z^*(x)) - \int_0^{x+z^*(x)} w''(x + z^*(x) - \xi)\, \varphi(\xi)\, d\xi - v''(z^*(x)) \int_{x+z^*(x)}^{\infty} \varphi(\xi)\, d\xi \right\}} \tag{4.9}$$

$$\underline{\text{for}}\ x < x_0 .$$

We see that, for $x < x_0$, dz^*/dx is negative and bounded by 1 in absolute value:

$$- 1 \leqq \frac{dz^*}{dx} < 0 , \qquad x < x_0 . \tag{4.10}$$

Hence, z^* is monotone decreasing in x. From (4.10) we get

$$\frac{d}{dx} [x + z^*(x)] \geqq 0 , \tag{4.11}$$

and this is valid for all x. From (4.9), we see also that if $v''(\eta) < 0$ for $\eta > 0$, then the strict inequality signs hold in (4.10) and (4.11). The result (4.11) tells us that $x + z^*(x)$ is monotone nondecreasing. Let us define

$$\rho = \lim_{x \to -\infty} [x + z^*(x)] . \tag{4.12}$$

By (3.26) we have that $\rho \geqq 0$. The determining equation for ρ is obtained from (3.22) by letting $x \to - \infty$, and noting that in this case $z^*(x) \to + \infty$, by virtue of (3.26). We get:

$$c - v'_{\infty} \int_{\rho}^{\infty} \varphi(\xi)\, d\xi - \int_0^{\rho} w'(\rho - \xi)\, \varphi(\xi)\, d\xi = 0 . \tag{4.13}$$

Since $c > 0$ we see that, in fact, $\rho > 0$.

If we now consider $x \to - \infty$ in (4.7) we get

$$\lim_{x \to -\infty} H'(x) = v'_{\infty} - c . \tag{4.14}$$

From (2.6) and the fact that $w'_\infty \geqq 0$ we have then:

$$\lim_{x \to -\infty} H'(x) > 0 \geqq \lim_{x \to \infty} H'(x) . \tag{4.15}$$

$H'(x)$ actually changes sign as x runs from $-\infty$ to $+\infty$. In fact there is a sign change between $-\infty$ and x_0, as we see from the fact that

$$H'(x_0) = - \int_0^{x_0} w'(x_0 - \xi)\, \varphi(\xi)\, d\xi < 0 . \tag{4.16}$$

We now calculate the second derivative of H; from (4.7) we get:

$$\begin{aligned} H''(x) = & \left\{ [v'(z^*(x)) - w'(0)]\, \varphi(x + z^*(x)) \right. \\ & \left. - \int_0^{x+z^*(x)} w''(x + z^*(x) - \xi)\, \varphi(\xi)\, d\xi \right\} \left(1 + \frac{dz^*}{dx}\right) \\ & - \int_0^{x+z^*(x)} v''(\xi - x)\, \varphi(\xi)\, d\xi \end{aligned} \tag{4.17}$$

for $x < 0$,

and

$$\begin{aligned} H''(x) = & \left\{ [v'(z^*(x)) - w'(0)]\, \varphi(x + z^*(x)) \right. \\ & \left. - \int_0^{x+z^*(x)} w''(x + z^*(x) - \xi)\, \varphi(\xi)\, d\xi \right\} \left(1 + \frac{dz^*}{dx}\right) \\ & - v'(0)\, \varphi(x) - \int_x^{x+z^*(x)} v''(\xi - x)\, \varphi(\xi)\, d\xi \end{aligned} \tag{4.18}$$

for $0 < x < x_0$,

and

$$H''(x) = - w'(0)\, \varphi(x) - \int_0^x w''(x - \xi)\, \varphi(\xi)\, d\xi \tag{4.19}$$

for $x > x_0$.

The information just gathered, concerning the function H, will suffice to investigate the Function J.

STUDY OF THE FUNCTION J(x)

We have

$$J(x) = L(x) + H(x) \tag{5.1}$$

and

$$L(x) = \begin{cases} \int_0^x h(x-\xi)\,\varphi(\xi)\,d\xi + \int_x^\infty p(\xi - x)\,\varphi(\xi)\,d\xi, & x > 0, \\ \int_0^\infty p(\xi - x)\,\varphi(\xi)\,d\xi, & x \leqq 0. \end{cases} \tag{5.2}$$

H and L are continuous functions, and therefore J is continuous. From (5.2) we calculate:

$$L'(x) = \begin{cases} \int_0^x h(x-\xi)\,\varphi(\xi)\,d\xi - \int_x^\infty p'(\xi - x)\,\varphi(\xi)\,d\xi, & x > 0, \\ -\int_0^\infty p'(\xi - x)\,\varphi(\xi)\,d\xi, & x < 0. \end{cases} \tag{5.3}$$

From this we see that also L' is continuous. We have already seen that H' is continuous. Hence, J' is continuous.

Differentiating once again, we get

$$L''(x) = \begin{cases} [h'(0) + p'(0)]\,\varphi(x) + \int_0^x h''(x-\xi)\,\varphi(\xi)\,d\xi \\ \qquad + \int_x^\infty p''(\xi - x)\,\varphi(\xi)\,d\xi, \quad x > 0, \\ \int_0^\infty p''(\xi - x)\,\varphi(\xi)\,d\xi, \quad x < 0. \end{cases} \tag{5.4}$$

Combining this with (4.17)—(4.19) we get

$$(5.5)\qquad J''(x) = \left\{ [v'(z^*(x)) - w'(0)]\,\varphi(x + z^*(x)) \right.$$

$$\left. \int_0^{x+z^*(x)} w''(x + z^*(x) - \xi)\,\varphi(\xi)\,d\xi \right\} \left(1 + \frac{dz^*}{dx}\right)$$

$$- \int_0^{x+z^*(x)} v''(\xi - x)\,\varphi(\xi)\,d\xi + \int_0^{\infty} p''(\xi - x)\,\varphi(\xi)\,d\xi$$

$$\text{for } x < 0$$

and

$$(5.6)\qquad J''(x) = \left\{ [v'(z^*(x)) - w'(0)]\,\varphi(x + z^*(x)) \right.$$

$$\left. - \int_0^{x+z^*(x)} w''(x + z^*(x) - \xi)\,\varphi(\xi)\,d\xi \right\} \left(1 + \frac{dz^*}{dx}\right)$$

$$- \int_x^{x+z^*(x)} v''(\xi - x)\,\varphi(\xi)\,d\xi + [h'(0) + p'(0) - v'(0)]\,\varphi(x)$$

$$+ \int_0^{x} h''(x - \xi)\,\varphi(\xi)\,d\xi + \int_x^{\infty} p''(\xi - x)\,\varphi(\xi)\,d\xi$$

$$\text{for } 0 < x < x_0$$

and

$$(5.7)\qquad J''(x) = -\int_0^{x} w''(x - \xi)\,\varphi(\xi)\,d\xi + [h'(0) + p'(0) - w'(0)]\,\varphi(x)$$

$$+ \int_0^{x} h''(x - \xi)\,\varphi(\xi)\,d\xi + \int_x^{\infty} p''(\xi - x)\,\varphi(\xi)\,d\xi$$

$$\text{for } x > x_0 .$$

We see immediately, by (2.1), (2.2), (2.4), (2.9) and (4.11), that J'' is nonnegative for $x < 0$. Taking account as well of (2.3) and (2.7), we see that J'' is strictly positive for $0 < x < x_0$. And by virtue of these also, noting that

(5.8) $$p'(0) > v'(0) > w'(0) ,$$

we see that, finally, J'' is strictly positive for $x > x_0$. Hence, we have the result that J is a convex function.

Let

(5.9) $$p'_\infty \stackrel{\text{def}}{=} \lim_{\eta \to \infty} p'(\eta) .$$

(p'_∞ may be infinite.) From (5.3) we see that

(5.10) $$\lim_{x \to -\infty} L'(x) = - p'_\infty .$$

Combining this with (4.14), we have

(5.11) $$\lim_{x \to -\infty} J'(x) = v'_\infty - p'_\infty - c .$$

From the properties of v and p, it follows by (2.7) that

(5.12) $$v'_\infty < p'_\infty .$$

Hence,

(5.13) $$\lim_{x \to -\infty} J'(x) < 0 .$$

Again from (5.3),

(5.14) $$\lim_{x \to \infty} L'(x) = h'_\infty .$$

Combining this with (4.8), we have

(5.15) $$\lim_{x \to \infty} J'(x) = h'_\infty - w'_\infty .$$

Thus, by (2.8),

(5.16) $$\lim_{x \to \infty} J'(x) > 0 .$$

Hence, the function J is U-shaped.

Now we are at the point of investigating J with regard to the question centering around (2.19). For this purpose we state, without proof, the following

LEMMA: Let $g(x)$ be defined for all $x \in (-\infty, \infty)$, and let it have the properties

$$\text{(5.17)} \qquad \begin{cases} \lim_{x \to -\infty} g'(x) < 0 < \lim_{x \to +\infty} g'(x), \\ g''(x) \geqq 0. \end{cases}$$

Let α and β be positive numbers, and define

$$\text{(5.18)} \qquad \psi(x) = g(x + \alpha) + \beta.$$

Finally, define

$$\text{(5.19)} \qquad \nu = - \lim_{x \to -\infty} g'(x).$$

Then:

(i) If the graphs of g and ψ do not intersect—in which case the graph of ψ lies entirely above that of g—then ν is necessarily finite, and in fact, $\nu \leqq \beta/\alpha$. Consequently, if $\nu = \infty$, then the graphs of g and ψ necessarily intersect in at least one point.

(ii) If the graphs of g and ψ intersect in a single point, and γ is the abscissa of this point, then $\psi(x) < g(x)$ for $x < \gamma$ and $\psi(x) > g(x)$ for $x > \gamma$.

(iii) If the graphs of g and ψ intersect in more than one point, then their common set is a straight line segment of slope $-\beta/\alpha$. If γ' and γ'' are the left and right end-points of the abscissa set of this common segment, then $\gamma'' < \infty$ necessarily; on the other hand, $\gamma' > -\infty$ if and only if $\nu > \beta/\alpha$. In this case of a common segment we have $\psi(x) < g(x)$ for $x < \gamma'$ and $\psi(x) > g(x)$ for $x > \gamma''$.

(iv) If $\nu < \infty$ and g has an asymptote as $x \to -\infty$, then:

(a) the graphs of g and ψ intersect if $\beta/\alpha < \nu$;

(b) if the graphs of g and ψ intersect, then $\beta/\alpha \leqq \nu$; thus, the graphs of g and ψ fail to intersect for all $\beta > \alpha\,\nu$;

(c) if $\beta/\alpha = \nu$, then either the graphs of g and ψ do not intersect or they coincide for all $x \leqq$ some γ'', according as the graph of g does not or does coincide with its asymptote to the left of some point.

We apply this lemma to our function J. From (5.11) we have that, since v'_∞ is finite, the question of whether or not $\nu < \infty$ obtains in our case depends on p'_∞. It follows immediately from the lemma, then, that, whatever be m and c_o, there is a γ' satisfying (2.19) if $p'_\infty = \infty$.

On the other hand, if $p'_\infty < \infty$ and $J(x)$ has an asymptote as $x \to -\infty$, then the graphs of $J(x)$ and $J(x + m) + c_o m$ will intersect for all sufficiently small positive values of $c_o m$, and will fail to intersect for all values of $c_o m$ greater than the pertinent critical value. This critical value depends only on the slope of the asymptote of J, which is given by (5.11). However, for completeness, we shall determine the asymptote of J fully explicitly.

To investigate this asymptote we shall make the convenient (and reasonable) assumption:

ASSUMPTION:

$$\mu \equiv \int_0^\infty \xi\, \varphi(\xi)\, d\xi < \infty ; \tag{5.20}$$

that is, that the demand variable has finite expectation.

If g is a function which has an asymptote as $x \to -\infty$, then the vertical intercept of the asymptote is given by

$$\lim_{x \to -\infty} [g(x) - xg'(x)] . \tag{5.21}$$

The same formula applies for $x \to +\infty$ if there is an asymptote in this direction. In our present case, we shall see that H has an asymptote as $x \to -\infty$ provided the function v has one as $x \to +\infty$, and that L has one as $x \to -\infty$ provided p does as $x \to +\infty$. J is the sum of L and H, and the Formula (5.21) is linear. Hence the vertical intercept of the asymptote of J is the sum of those for L and H. Let λ_L and λ_H denote the latter two vertical intercepts, respectively. We first calculate λ_L.

We suppose that the function p has an asymptote as $x \to \infty$. Let λ_p be the vertical intercept of this asymptote. Applying (5.21) to (5.2) and (5.3) we have:

$$\lambda_L = \lim_{x \to -\infty} \left[\int_0^\infty p(\xi - x)\, \varphi(\xi)\, d\xi + x \int_0^\infty p'(\xi - x)\, \varphi(\xi)\, d\xi \right] \tag{5.22}$$

$$= \lim_{x \to -\infty} \int_0^\infty [p(\xi - x) + xp'(\xi - x)]\, \varphi(\xi)\, d\xi$$

$$= \lim_{x \to -\infty} \int_0^\infty \{[p(\xi - x) - (\xi - x)\, p'(\xi - x)] + \xi\, p'(\xi - x)\}\, \varphi(\xi)\, d\xi$$

$$= \int_0^\infty \{\lambda_p + p'_\infty\, \xi\}\, \varphi(\xi)\, d\xi ,$$

and so

$$\lambda_L = \lambda_p + p'_\infty\, \mu . \tag{5.23}$$

Next, we calculate λ_H. Utilizing (2.14), (3.1), (3.2) and (4.7), we have, for $x < 0$,

(5.24) $$H(x) - xH'(x) = cz^*(x)$$

$$- \int_0^{x+z^*(x)} v(\xi - x)\,\varphi(\xi)\,d\xi - v(z^*(x)) \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi$$

$$- \int_0^{x+z^*(x)} w(x + z^*(x) - \xi)\,\varphi(\xi)\,d\xi$$

$$- x \int_0^{x+z^*(x)} v'(\xi - x)\,\varphi(\xi)\,d\xi - xv'(z^*(x)) \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi$$

$$+ cx$$

$$= c(x + z^*(x)) - \int_0^{x+z^*(x)} [v(\xi - x) + xv'(\xi - x)]\,\varphi(\xi)\,d\xi$$

$$- [v(z^*(x)) + xv'(z^*(x))] \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi$$

$$- \int_0^{x+z^*(x)} w(x + z^*(x) - \xi)\,\varphi(\xi)\,d\xi$$

$$= c \cdot (x + z^*(x)) - \int_0^{x+z^*(x)} [v(\xi - x) - (\xi - x)\,v'(\xi - x)]\,\varphi(\xi)\,d\xi$$

$$- \int_0^{x+z^*(x)} \xi\,v'(\xi - x)\,\varphi(\xi)\,d\xi$$

$$- [v(z^*(x)) - z^*(x)\,v'(z^*(x))] \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi$$

$$- (x + z^*(x))\,v'(z^*(x)) \int_{x+z^*(x)}^{\infty} \varphi(\xi)\,d\xi$$

$$- \int_0^{x+z^*(x)} w(x + z^*(x) - \xi)\,\varphi(\xi)\,d\xi\ .$$

In this last form, we see that every term has a limit, provided the function $v(\eta)$ has an asymptote as $\eta \to \infty$. We assume that this is so, and we denote the vertical intercept of this asymptote by λ_v. Then, taking limits, we get:

$$\lambda_H = c\rho - \lambda_v \int_0^\rho \varphi(\xi)\,d\xi - v'_\infty \int_0^\rho \xi\,\varphi(\xi)\,d\xi \tag{5.25}$$

$$- \lambda_v \int_\rho^\infty \varphi(\xi)\,d\xi - pv'_\infty \int_\rho^\infty \varphi(\xi)\,d\xi - \int_0^\rho w(\rho - \xi)\,\varphi(\xi)\,d\xi ,$$

or, simplifying,

$$\lambda_H = c\rho - \lambda_v - pv'_\infty \int_\rho^\infty \varphi(\xi)\,d\xi - v'_\infty \int_0^\rho \xi\,\varphi(\xi)\,d\xi - \int_0^\rho w(\rho - \xi)\,\varphi(\xi)\,d\xi . \tag{5.26}$$

And now combining (5.23) and (5.26) we have

$$\lambda_J = (\lambda_p - \lambda_v) + p'_\infty \mu - v'_\infty \int_0^\rho \xi\,\varphi(\xi)\,d\xi \tag{5.27}$$

$$+ c\rho - pv'_\infty \int_\rho^\infty \varphi(\xi)\,d\xi - \int_0^\rho w(\rho - \xi)\,\varphi(\xi)\,d\xi .$$

Thus, finally, we have, by (5.11) and with λ_J as given by (5.27), the following equation for the asymptote of J as $x \to -\infty$:

$$y = (v'_\infty - p'_\infty - c)\,x + \lambda_J . \tag{5.28}$$

Now, from (5.28) we are able to calculate that the asymptote of $J(x + m)$ is

$$y = (v'_\infty - p'_\infty - c)\,x + [m(v'_\infty - p'_\infty - c) + \lambda_J] . \tag{5.29}$$

And we see that the line (5.29) is at a vertical distance

$$m(c + p'_\infty - v'_\infty) \tag{5.30}$$

below the line (5.28). (Notice that this quantity (5.30) does not depend on λ_J, and could have been reasoned to without explicit knowledge of λ_J.)

The quantity (5.30) is the critical value to which we referred at the beginning of our discussion of the asymptote of J. Thus, if $c_o m$ is less than (5.30), then the graphs of $J(x)$ and $J(x + m) + c_o m$ will intersect; on the other hand, if $c_o m$ is greater than (5.30), then

these two graphs will not intersect. In the case of intersection, the abscissa of any point of intersection is a number γ' fulfilling (2.19). Thus, we may state:

Proposition II

If (in the case of existence of asymptotes)

$$c_0 < c + p'_\infty - v'_\infty , \tag{5.31}$$

then there is a number γ' such that (2.19) holds. This γ' is the abscissa of any one of the points of intersection of the graphs of $J(x)$ and $J(x + m) + c_0 m$.

If

$$c_0 > c + p'_\infty - v'_\infty , \tag{5.32}$$

then the graph of $J(x + m) + c_0 m$ lies everywhere above the graph of $J(x)$, and no emergency level is to be defined. We state also formally the additional result noted above:

Proposition III

If $p'_\infty = \infty$, then there is a γ' such that (2.19) holds. Again, this number is the abscissa of any one of the points of intersection of the graphs of $J(x)$ and $J(x + m) + c_0 m$.

CONCLUDING REMARKS

To round out our discussion it remains only to add an explicit word regarding the optimal ordering policy. According to the discussion in the first section, Formula (2.18) gives us the optimal cost function when the (or,an) optimal value is taken for γ, as discovered in the subsequent analysis and expressed by the Lemma and by Propositions II and III. An optimal γ may be finite or $-\infty$. If it is finite, then, of course, the optimal ordering procedure calls for placing the emergency order for m units when $x \leqq \gamma$. But furthermore, we see in this case that when $x > \gamma$ the optimal regular order is the quantity z which minimizes $G(z; x)$, and when $x \leqq \gamma$ the optimal regular order is the quantity z which minimizes $G(z; x + m)$. Thus, under the assumption (3.23), when $x > \gamma$ the optimum regular order z is the unique root of (3.22) as it stands, and when $x \leqq \gamma$ the optimum regular order z is the unique root of (3.22) with x replaced by $x + m$.

REFERENCE

[1] Arrow, Karlin, and Scarf, Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, 1958.

* * *

RECENT PUBLICATIONS

EDITOR'S NOTE: Listing of a publication in this section is for record and reference only and does not constitute an endorsement of point of view or advocacy of use.

PROBABILITY AND STATISTICS, edited by Ulf Grenander, John Wiley and Sons, New York, 1959, 434 pp.

This book is a collection of studies broadly covering statistics and probability. The articles range from the very practical such as Tukey's engineering approach to time series, to the very theoretical such as Masani's and Wiener's analysis of nonlinear prediction. The articles tend to be expository and heuristic, giving the nonspecialist a very readable survey of other fields.

Time Series is given good coverage, with generous bibliographies, and emphasis on the spectrum. Bartlett gives a general picture. Tukey gives practical suggestions for measuring the spectrum. Rosenblatt treats both the regression coefficients and the spectrum of the residual process. Grenander outlines some nonlinear problems. Wiener and Masani develop nonlinear predictions in terms of moments, and Masani writes a further article on the Wold decomposition and monotone matrix valued functions.

For the social scientists there are three articles: "Some Scaling Models and Estimation Procedures in the Latent Class Model," by T. W. Anderson; "A Survey of Results in the Collective Theory of Risk," by C. O. Segerdahl; "Ends and Means in Econometric Model Building," a very complete survey by Wold.

In probability theory, there are several delightful facets: "On Combinatorial Methods in Fluctuation Theory," by Feller; "Some Remarks on Stable Processes with Independent Increments," by Kac; "Unitary Dilations of Markov Transition Operators, and the Corresponding Integral Representations for Transition-Probability Matrices," by D. G. Kendall; "A Markov Chain Theorem," by Doob; and "Construction du Processus de W. Feller et H. P. McKean en Partant du Mouvement Brownien," by Levy.

The remaining articles cover sundry topics in Statistics: "Design of Linear Experiments," by Elfving; "The Restricted Chi-Square Test," by Fix, Hodges, and Lehman; "Optimal Asymptotic Tests of Composite Statistical Hypotheses," by Neyman; "Sequential Estimation of the Mean of a Normal Population," by H. Robbins; and "Nonparametric Statistical Inference," by Wilks.

It was a disappointment that in this volume, dedicated to Harold Cramer, there was no resumé of his great works. Such a theme might also have given it a little cohesion, which it rather lacks.

(Reviewed by Thomas Wonnacott)

LINEAR PROGRAMMING IN INDUSTRY, THEORY AND APPLICATIONS, AN INTRODUCTION, by Sven Danø, Springer-Verlag, Vienna, 1960, 120 pp.

From the Preface, the twofold purpose of the book is to (1) "provide a university text of Linear Programming for students of economics or operations research," and (2) for use by "engineers, business executives, managers, and others responsible for the organization and planning of industrial operations."

The development is quite elementary, and as such can reach a wide variety of readers. Although the full scope of linear programming application in industry can hardly be confined to so few pages, the author appears to have recognized and to have imparted this thought.

There are nine Chapters, each brief, wherein the author covers the simplex method, some of the usually employed industrial examples, and some peripheral topics such as degeneracy, alternate optima, short cuts, duality interpretation, and variation of coefficients.

At one or two points, there seems to be some confusion in the author's mind as to what constitutes "mathematics." From the Preface, "the mathematical exposition is purely elementary, involving only simple linear relations," ..., "as compared with the use of matrix algebra, amply justifying the sacrifice of mathematical elegance and typographical simplicity in explaining the simplex method."

On p. 77, after exposing the simplex method: "No mathematical operations in the proper sense of the word are involved, ..."

Indeed, in the opinion of the reviewer, the use of this "elegant mathematics" would be unnecessary and out of place in this book.

On p. 39 (and elsewhere) the use of numbers, xj, v_i as "other things" ("final product x_j," "raw material v_i") may cause confusion.

The "Fundamental Theorem" states that a linear functional takes on its optimum at an extreme point of a convex polyhedron (Chapter 2).

There is no material, except possibly for some examples, which is not already in book form. But what material is included, is presented in clear logical order, with no irrelevant deviations, so that a good introduction to the applicability of linear programming results.

(Reviewed by C. E. Lemke)

MARKOV LEARNING MODELS FOR MULTIPERSON INTERACTIONS, by Patrick Suppes and Richard C. Atkinson, Stanford University Press, Stanford, California, 1960, 296 pp., $8.25.

The finding presented in this volume represent an intimate interaction of theory and experiment that is unusual in the history of psychology, and almost unique in the field of social behavior. The methodology is close in spirit, though not in specific mathematical technique, to that involved in applying the laws of mechanics to the two-body problem. A particular case of mathematical learning theory which is known to describe the behavior of the individual subject in certain standard learning experiments is applied to a series of experiments concerned with the interaction of two individuals over a series of trials in situations akin to simple games of strategy.

With respect to mathematical tractability and fertility in generating new methods of data analysis, the theory developed by Suppes and Atkinson represents a distinct advance over

hitherto published approaches to the same type of problem. The key to this advance is their conception of the system of two interacting individuals as a Markov chain. By appropriate interpretations of the axioms of statistical learning theory, they show that all information about the past learning history of the individual in a given situation which is (within the theory) relevant to prediction of their joint behavior on a particular trial can be subsumed in the "state" of the system at the beginning of that trial. Then, using the well-developed methodology of Markov chains, the authors achieve considerable success in predicting asymptotic choice probabilities for a variety of simple game situations, and make a substantial beginning on the more difficult task of accounting for the detailed sequential structure of the data. A useful byproduct of this work is a considerable body of new or improved techniques for generating theorems and estimating parameters in the general class of stimulus sampling theories.

(Reviewed by W. K. Estes)

ADAPTIVE CONTROL PROCESSES—A GUIDED TOUR, by Richard Bellman, Princeton University Press, Princeton, New Jersey, 1961, 255 pp., $6.50.

This excellent book is oriented toward practicing scientists from all disciplines concerned with control of complex systems; specifically included are not only mathematicians and electrical engineers but also psychologists, biologists, operations analysts, and mathematical economists (e.g., one discussion concerns optimal inventory problems in which the distribution of demand must be learned as the process goes on). However, it should not be inferred from the title that this volume is simply a general review of the various techniques applicable to adaptive control processes. It is, rather, a treatise on dynamic programming, with contrasts drawn to other approaches. This "guided tour" to adaptive control might be likened to a tour of the battlefield at Gettysburg conducted by a Southerner; the tourist is shown all the important points, but the accompanying lecture is not exactly unbiased. Specifically, the many accomplishments already achieved with traditional techniques are largely glossed over in this book, and the reader must get to page 200 before he is informed that simpler mathematical techniques (than dynamic programming) may in fact satisfy his needs.

The book begins with a review of feedback control and its conventional treatment via the calculus of variations; this part includes an excellent denumeration of the assumptions and difficulties involved in the conventional approaches. Dynamic programming is then introduced as an alternative (and more powerful) concept, and various implications and modifications are discussed at some length (computational aspects, two-point boundary value problems, logical systems, random processes, theory of games, and the like). Finally, in the last quarter of the book, adaptive processes are introduced and treated by removing the various problem constraints previously required; Bellman limits adaptive control processes (quite properly) to those which involve some form of learning about the system under consideration.

Presentation throughout the book is exceptionally clear. The reader is not required to memorize every technique and symbol as he goes along in order to understand subsequent arguments. Each chapter starts with a brief introduction to orient the reader, and specific points are then developed logically, amplified with appropriate comments and discussion, and exemplified with problems drawn from a wide variety of disciplines. The chapters conclude with excellent annotated bibliographies. This reviewer enjoyed also the delightful snatches of

poetry and prose which head each chapter. In short, despite the "hard sell" of dynamic programming, this book should be an interesting and useful addition to the library of any mathematically inclined scientist concerned with complex systems.

(Reviewed by R. H. Wilcox)

* * *

U. S. GOVERNMENT PRINTING OFFICE : 1961 O - 617404

INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Washington 25, D. C. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 50 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.